Kybernetika

Peter Harremoës

Bounds on tail probabilities for negative binomial distributions

Kybernetika, Vol. 52 (2016), No. 6, 943–966

Persistent URL: http://dml.cz/dmlcz/145999

BOUNDS ON TAIL PROBABILITIES FOR NEGATIVE BINOMIAL DISTRIBUTIONS

Peter Harremoës

In this paper we derive various bounds on tail probabilities of distributions for which the generated exponential family has a linear or quadratic variance function. The main result is an inequality relating the signed log-likelihood of a negative binomial distribution with the signed log-likelihood of a Gamma distribution. This bound leads to a new bound on the signed log-likelihood of a binomial distribution compared with a Poisson distribution that can be used to prove an intersection property of the signed log-likelihood of a binomial distribution compared with a standard Gaussian distribution. All the derived inequalities are related and they are all of a qualitative nature that can be formulated via stochastic domination or a certain intersection property.

Keywords: tail probability, exponential family, signed log-likelihood, variance function, inequalities

Classification: 60E15, 62E17, 60F10

1. INTRODUCTION

Let $X_1, \ldots, X_n$ be i.i.d. random variables such that the moment generating function $\beta \curvearrowright E[\exp(\beta X_1)]$ is finite in a neighborhood of zero. For a fixed value of x one is interested in approximating the tail distribution: $\Pr(\sum_{i=1}^{n} X_i \leq n \cdot x)$. If x is close to the mean of X_1 one would usually approximate the tail probability by the tail probability of a Gaussian random variable. If x is far from the mean of X_1 the tail probability can be estimated using large deviation theory. According to the Sanov theorem the probability that the deviation from the mean is as large as x is of the order $\exp(-D)$ where D is a constant that can be calculated as an information divergence between two distributions in an exponential family. The more precise formulation of the result is that

$$-\frac{\ln\left(\Pr(\sum_{i=1}^{n} X_i \leq n \cdot x)\right)}{n} \to D$$

for $n \to \infty$. Bahadur and Rao [2, 3] improved the estimate of this large deviation probability, and in [5] such Gaussian tail approximations were extended to situations where one normally uses large deviation techniques.

DOI: 10.14736/kyb-2016-6-0943

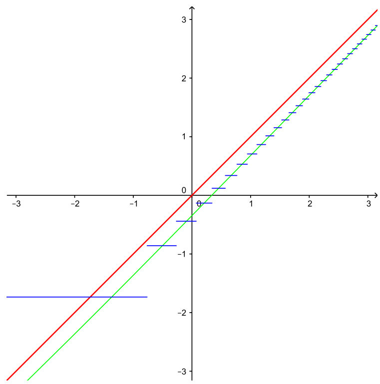

Fig. 1. Plot of the quantiles of a standard Gaussian vs. the quantiles of the signed log-likelihood of the negative binomial distribution $neg\,(1, 3.5)$ (horisontal steps) and of the signed log-likelihood of the Gamma distribution $\Gamma\,(1, 3.5)$ (lower full line). The line through (0,0) corresponds to a perfect mach with a Gaussian.

The distribution of the signed log-likelihood is close to a standard Gaussian for a variety of distributions. Asymptotic results for large sample sizes are not new [2, 3], but in this paper we are interested in inequalities that hold for any sample size. Some inequalities of this type can be found in [1, 7, 6, 10, 11]. In [6] a tail probability of the log-likelihood of a negative binomial distribution was compared with the tail probability of a standard Gaussian distribution. The result can be visualized by Figure 1 where the quantiles of the signed log-likelihood of a negative binomial distribution (blue) are plotted against the corresponding quantiles of a standard Gaussian. The result in [6] is that the right end points of the horizontal lines are to the right of the red line that corresponds a perfect match with a Gaussian distribution. In [6] there is no result related to the left end points of the blue lines and Figure 1 demonstrates that the left end points can be above or below the red line. In Figure 1 the green curve depicts the log-likelihood of a Gamma distribution against a standard Gaussian and we see that the green curve intersects all the horizontal lines. This reflects that the negative binomial distributions and the Gamma distributions are discrete and continuous versions of waiting times of the same type of process. We will prove the intersection property and use it to derive a new inequality relation binomial and Poisson distributions.

In this paper we let τ denote the circle constant 2π and ϕ will denote the standard Gaussian density

$$\frac{\exp\left(-\frac{x^2}{2}\right)}{\tau^{1/2}}.$$

We let Φ denote the distribution function of the standard Gaussian

$$\Phi\,(t) = \int_{-\infty}^{t} \phi\,(x)\;\mathrm{d}x\ .$$

The rest of the paper is organized as follows. In Section 2 we define the signed log-likelihood of an exponential family and look at some of the fundamental properties of the signed log-likelihood. The proof of the main result concerning negative binomial distributions is quite long. Therefore as much material as possible has been moved from the main proof to some sections with some more preliminary results on the exponential distributions (Section 3) and more general Gamma distributions (Section 4), and geometric distributions (Section 5) and then we generalize the results to negative binomial distributions (Section 6). The negative binomial distributions are waiting times in Bernoulli processes, so in Section 7 our inequalities between negative binomial distributions and Gamma distributions are translated into inequalities between binomial distributions and Poisson distributions. Combined with our domination inequalities for Gamma distributions we obtain an intersection inequality between binomial distributions and the standard Gaussian distribution. In this paper the focus is on intersection inequalities and stochastic domination inequalities, but in the discussion we mention some related inequalities of other types and how our inequalities might be tightened.

2. THE SIGNED LOG-LIKELIHOOD FOR EXPONENTIAL FAMILIES

Let P_0 denote a probability measure on the real numbers. For any real number β the *moment generating function* is given by $\mathrm{Z}(\beta) = \int \exp(\beta \cdot x)\, \mathrm{d}P_0 x$. When $Z(\beta) < \infty$ the distributions P_β are given by

$$\frac{\mathrm{d}P_\beta}{\mathrm{d}P_0}(x) = \frac{\exp(\beta \cdot x)}{Z(\beta)}$$

and these distributions form a one-dimensional exponential family. Let P^μ denote the element in the exponential family with mean value μ, and let $\hat{\beta}(\mu)$ denote the corresponding maximum likelihood estimate of β. Let μ_0 denote the mean value of P_0. Then

$$D(P^\mu \| P_0) = \int \ln\left(\frac{\mathrm{d}P^\mu}{\mathrm{d}P_0}(x)\right) \mathrm{d}P^\mu x.$$

With this definition the divergence D becomes a differentiable function of μ. The variance function of an exponential family is defined so that $V(\mu)$ is the variance of P^μ. The variance functions uniquely characterizes the corresponding exponential families and the most important exponential families have very simple variance functions. If we know the variance function the divergence can be calculated according to the following formula.

$$D(P^{\mu_1} \| P^{\mu_2}) = \int_{\mu_1}^{\mu_2} \frac{\mu - \mu_1}{V(\mu)}\, \mathrm{d}\mu.$$

Definition 2.1. (From Barndorff-Nielsen [4]) Let X be a random variable with distribution P_0. Then the *signed log-likelihood* $G(X)$ of X is the random variable given by

$$G(x) = \begin{cases} -\left[2D(P^x \| P_0)\right]^{1/2}, & \text{for } x < \mu_0; \\ +\left[2D(P^x \| P_0)\right]^{1/2}, & \text{for } x \geq \mu_0. \end{cases}$$

We will need the following general lemma.

Lemma 2.2. If the variance function is increasing then

$$\frac{G(x)}{x-\mu_0}$$

is a decreasing function of x.

Proof. We have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{G(x)}{x-\mu_0}\right) &= \frac{(x-\mu_0)\frac{D'(x)}{G(x)} - G(x)}{(x-\mu_0)^2} \\ &= \frac{(x-\mu_0)\int_x^{\mu_0}\frac{-1}{V(\mu)}\,\mathrm{d}\mu - 2D}{(x-\mu_0)^2 G(x)} \\ &= \frac{(x-\mu_0)\int_{\mu_0}^{x}\frac{1}{V(\mu)}\,\mathrm{d}\mu - 2D}{(x-\mu_0)^2 G(x)}.\end{aligned}$$

We have to prove that the numerator is positive for $x<\mu_0$ and negative for $x>\mu_0$. The numerator can be calculated as

$$\begin{aligned}(x-\mu_0)\int_{\mu_0}^{x}\frac{1}{V(\mu)}\,\mathrm{d}\mu - 2D &= (x-\mu_0)\int_{\mu_0}^{x}\frac{1}{V(\mu)}\,\mathrm{d}\mu + 2\int_{\mu_0}^{x}\frac{\mu-x}{V(\mu)}\,\mathrm{d}\mu \\ &= \int_{\mu_0}^{x}\left(\frac{x-\mu_0}{V(\mu)} + 2\frac{\mu-x}{V(\mu)}\right)\mathrm{d}\mu \\ &= \int_{\mu_0}^{x}\frac{2\mu-\mu_0-x}{V(\mu)}\,\mathrm{d}\mu.\end{aligned}$$

If $x>\mu_0$ then

$$\begin{aligned}\int_{\mu_0}^{x}\frac{2\mu-\mu_0-x}{V(\mu)}\,\mathrm{d}\mu &= \int_{\mu_0}^{\frac{x+\mu_0}{2}}\frac{2\mu-\mu_0-x}{V(\mu)}\,\mathrm{d}\mu + \int_{\frac{x+\mu_0}{2}{}_0}^{x}\frac{2\mu-\mu_0-x}{V(\mu)}\,\mathrm{d}\mu \\ &\le \int_{\mu_0}^{\frac{x+\mu_0}{2}}\frac{2\mu-\mu_0-x}{V\left(\frac{x+\mu_0}{2}\right)}\,\mathrm{d}\mu + \int_{\frac{x+\mu_0}{2}{}_0}^{x}\frac{2\mu-\mu_0-x}{V\left(\frac{x+\mu_0}{2}\right)}\,\mathrm{d}\mu \\ &= \int_{\mu_0}^{x}\frac{2\mu-\mu_0-x}{V\left(\frac{x+\mu_0}{2}\right)}\,\mathrm{d}\mu = 0.\end{aligned}$$

The inequality for $x<\mu_0$ is proved in the same way. □

3. EXPONENTIAL DISTRIBUTIONS

Although the tail probabilities of the exponential distribution are easy to calculate the inequalities related to the signed log-likelihood of the exponential distribution are non-trivial and will be useful later.

The exponential distribution Exp^{θ} has density

$$f(x) = \frac{1}{\theta}\exp\left(-\frac{x}{\theta}\right),\ x\ge 0.$$

The distribution function is

$$\Pr(X \le x) = \int_0^x \frac{1}{\theta} \exp\left(-\frac{t}{\theta}\right) \mathrm{d}t = 1 - \exp\left(-\frac{x}{\theta}\right),\ x \ge 0.$$

The mean of the exponential distribution Exp^θ is θ and the variance is θ^2 so the variance function is $V(\mu) = \mu^2$. The divergence can be calculated as

$$\begin{aligned} D\left(Exp^{\theta_1} \| Exp^{\theta_2}\right) &= \int_{\theta_1}^{\theta_2} \frac{\mu - \theta_1}{\mu^2} \mathrm{d}\mu \\ &= \frac{\theta_1}{\theta_2} - 1 - \ln \frac{\theta_1}{\theta_2}. \end{aligned}$$

This is the well-known Itakura-Saito divergence. We see that

$$\begin{aligned} G_{Exp^\theta}(x) &= \pm \left[2\left(\frac{x}{\theta} - 1 - \ln \frac{x}{\theta}\right)\right]^{1/2} \\ &= \gamma\left(\frac{x}{\theta}\right) \end{aligned}$$

where γ denotes the function

$$\gamma(x) = \begin{cases} -\left[2(x - 1 - \ln x)\right]^{1/2}, & \text{when } x \le 1; \\ +\left[2(x - 1 - \ln x)\right]^{1/2}, & \text{when } x > 1. \end{cases}$$

Note that the *saddle-point approximation* is exact for the family of exponential distributions, i. e.

$$f(x) = \frac{\tau^{1/2}}{\mathrm{e}} \cdot \frac{\phi(G(x))}{[V(x)]^{1/2}}.$$

Lemma 3.1. The density of the signed log-likelihood of an exponential random variable is given by

$$\frac{\tau^{1/2}}{\mathrm{e}} \cdot \frac{z\phi(z)}{\gamma^{-1}(z) - 1}.$$

Proof. Let X be a Exp^θ distributed random variable. Without loss of generality we may assume that $\theta = 1$. The density of the signed log-likelihood is

$$\begin{aligned} \frac{f\left(\gamma^{-1}(z)\right)}{\gamma'\left(\gamma^{-1}(z)\right)} &= \frac{\frac{\tau^{1/2}}{\mathrm{e}} \cdot \frac{\phi\left(\gamma\left(\gamma^{-1}(z)\right)\right)}{\left[V\left(\gamma^{-1}(z)\right)\right]^{1/2}}}{\gamma'\left(\gamma^{-1}(z)\right)} \\ &= \frac{\tau^{1/2}}{\mathrm{e}} \cdot \frac{\phi(z)}{\left[V\left(\gamma^{-1}(z)\right)\right]^{1/2} \gamma'\left(\gamma^{-1}(z)\right)}. \end{aligned}$$

The variance function is $V(x) = x^2$ so the density is

$$\frac{\tau^{1/2}}{\mathrm{e}} \cdot \frac{\phi(z)}{\gamma^{-1}(z) \cdot \gamma'\left(\gamma^{-1}(z)\right)}.$$

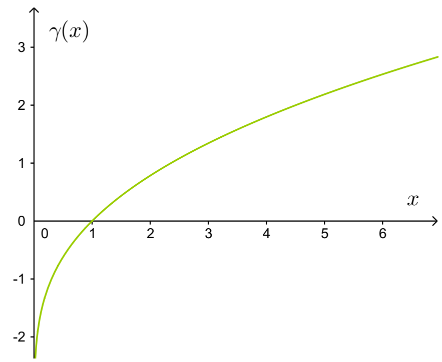

Fig. 2. The signed log-likelihood $\gamma(x)$ of an exponential distribution.

From $\gamma^2 = 2D$ it follows that $\gamma \cdot \gamma' = D'$ so that

$$\gamma'(z) = \frac{\frac{\mathrm{d}D}{\mathrm{d}z}}{\gamma(z)} = \frac{\frac{1}{\theta} - \frac{1}{z}}{\gamma(z)}.$$

Hence the density of $\gamma(X)$ can be written as

$$\frac{\tau^{1/2}}{\mathrm{e}} \cdot \frac{\phi(z)}{\gamma^{-1}(z) \cdot \frac{\frac{1}{\theta} - \frac{1}{\gamma^{-1}(z)}}{\gamma(\gamma^{-1}(z))}} = \frac{\tau^{1/2}}{\mathrm{e}} \cdot \frac{z\phi(z)}{\gamma^{-1}(z) - 1},$$

which proves the lemma. □

Lemma 3.2. (From Harremoës and Tusnády [7]) Let X_1 and X_2 denote random variables with density functions f_1 and f_2. If there exists a real number x_0 such that $f_1(x) \geq f_2(x)$ for $x \leq x_0$ and $f_1(x) \leq f_2(x)$ for $x \geq x_0$, then X_1 is stochastically dominated by X_2. In particular, if $\frac{f_2(x)}{f_1(x)}$ is increasing then X_1 is stochastically dominated by X_2.

Theorem 3.3. (From Harremoës and Tusnády [7]) The signed log-likelihood of an exponentially distributed random variable is stochastically dominated by the standard Gaussian.

The proof below is a simplified version of the proof in [7].

P r o o f. The quotient between the density of a standard Gaussian and the density of $G(X)$ is

$$\frac{\mathrm{e}}{\tau^{1/2}} \cdot \frac{\gamma^{-1}(z) - 1}{z}.$$

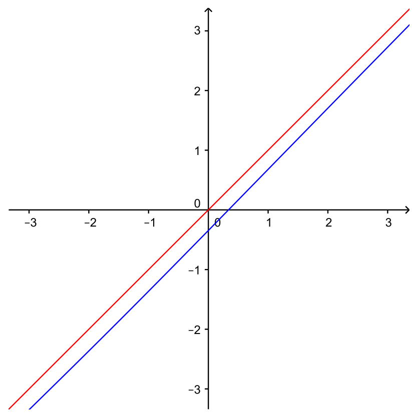

Fig. 3. Plot of the quantiles of a standard Gaussian vs. the quantiles of the signed log-likelihood of an exponential distribution.

We have to prove that this quotient is increasing. The function γ is increasing so it is sufficient to prove that $\frac{t-1}{\gamma(t)}$ is increasing or equivalently that

$$\frac{\gamma(t)}{t-1}$$

is decreasing. This follows from Lemma 2.2 because the variance function is increasing.

$\square$

4. GAMMA DISTRIBUTIONS

The sum of k exponentially distributed random variables is Gamma distributed $\Gamma(k,\theta)$ where k is called the shape parameter and θ is the scale parameter. It has density

$$f(x) = \frac{1}{\theta^k}\frac{1}{\Gamma(k)}x^{k-1}\exp\left(-\frac{x}{\theta}\right)$$

and this formula is used to define the Gamma distribution when k is not an integer. The mean of the Gamma distribution $\Gamma(k,\theta)$ is $\mu = k\cdot\theta$ and the variance is $k\cdot\theta^2$ so the variance function is $V(\mu) = \mu^2/k$. The divergence can be calculated as

$$\begin{aligned} D(\Gamma(k,\theta_1)\,\|\Gamma(k,\theta_2)) &= \int_{k\theta_1}^{k\theta_2}\frac{\mu-k\theta_1}{\mu^2/k}\,\mathrm{d}\mu \\ &= k\left(\frac{\theta_1}{\theta_2}-1-\ln\frac{\theta_1}{\theta_2}\right). \end{aligned}$$

Therefore we have that

$$G_{\Gamma(k,\theta)}(x) = k^{1/2}\gamma\left(\frac{x}{k\theta}\right).$$

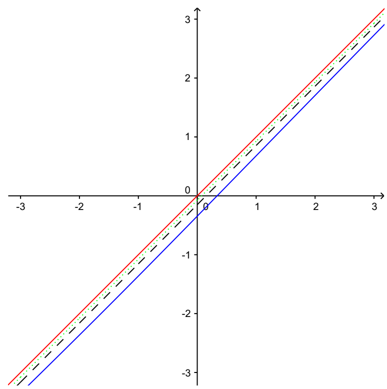

Fig. 4. The quantiles of a standard Gaussian vs. Gamma distributions for $k = 1$ (full), $k = 5$ (dash), and $k = 20$ (dot). The line through (0,0) corresponds to a perfect match with a standard Gaussian.

Note that the *saddle-point approximation* is exact for the family of Gamma distributions, i. e.

$$\begin{aligned} f(x) &= \frac{k^k \exp(-k)}{\Gamma(k)} \cdot \frac{\exp\left(-k\left(\frac{x}{k\theta} - 1 - \ln \frac{x}{k\theta}\right)\right)}{x} \\ &= \frac{k^k \tau^{1/2} \exp(-k)}{\Gamma(k)\, k^{1/2}} \cdot \frac{\phi\left(G_{\Gamma(k,\theta)}(x)\right)}{[V(x)]^{1/2}}. \end{aligned}$$

The following lemma is proved in the same way as Lemma 3.1.

Lemma 4.1. The density of the signed log-likelihood of a Gamma random variable is given by

$$\frac{k^k \tau^{1/2} \exp(-k)}{\Gamma(k)\, k^{1/2}} \cdot \frac{\frac{z}{k^{1/2}} \phi(z)}{\gamma^{-1}\left(\frac{z}{k^{1/2}}\right) - 1}.$$

Theorem 4.2. (From Harremoës and Tusnády [7]) The signed log-likelihood of a Gamma distributed random variable is stochastically dominated by the standard Gaussian, i. e.

$$\Pr(X \le x) \ge \Phi(G_\Gamma(x)).$$

P r o o f. This is proved in the same way as the corresponding result for exponential distributions. □

Theorem 4.3. Let X_1 and X_2 denote Gamma distributed random variables with shape parameters k_1 and k_2. Then the signed log-likelihood of X_1 is dominated by the signed log-likelihood of X_2 if and only if $k_1 \le k_2$.

P r o o f. We have to prove that

$$\frac{\frac{z}{k_1^{1/2}}\phi\left(z\right)}{\gamma^{-1}\left(\frac{z}{k_1^{1/2}}\right)-1} \leq \frac{\frac{z}{k_2^{1/2}}\phi\left(z\right)}{\gamma^{-1}\left(\frac{z}{k_2^{1/2}}\right)-1}$$

for $z > 0$ and the reverse inequality for $z < 0$. For $z > 0$ the inequality is equivalent to

$$\frac{\gamma^{-1}\left(\frac{z}{k_2^{1/2}}\right)-1}{\frac{z}{k_2^{1/2}}} \leq \frac{\gamma^{-1}\left(\frac{z}{k_1^{1/2}}\right)-1}{\frac{z}{k_1^{1/2}}}.$$

This follows because the function

$$\frac{t-1}{\gamma(t)}$$

is increasing. □

5. GEOMETRIC DISTRIBUTIONS

A *geometric distribution* can be obtained by compounding a Poisson distribution $Po\left(\lambda\right)$ with rate parameter λ distributed according to an exponential distribution $Exp\left(\theta\right)$. This geometric distribution will be denoted by Geo^{θ}. We note that this is an unusual way of parameterizing the geometric distributions, but it will be useful for some of our calculations. Since λ is both the mean and the variance of $Po\left(\lambda\right)$ the mean of Geo^{θ} is θ and the variance function is $V\left(\mu\right) = \mu + \mu^2$.

For $m = 0, 1, 2, \ldots$ the point probabilities of a geometric distribution can be written as

$$\begin{aligned}\Pr\left(M=m\right) &= \int_0^{\infty} \frac{\lambda^m}{m!}\exp\left(-\lambda\right)\cdot\frac{1}{\theta}\exp\left(-\frac{\lambda}{\theta}\right)\,\mathrm{d}\lambda \\ &= \int_0^{\infty}\frac{\left(\theta t\right)^m}{m!}\exp\left(-\theta t\right)\cdot\exp\left(-t\right)\,\mathrm{d}t \\ &= \frac{\theta^m}{\left(\theta+1\right)^{m+1}}.\end{aligned}$$

The distribution function can be calculated as

$$\begin{aligned}\Pr\left(M \leq m\right) &= \sum_{j=0}^{m}\frac{\theta^j}{\left(\theta+1\right)^{j+1}} \\ &= 1-\left(\frac{\theta}{\theta+1}\right)^{m+1}.\end{aligned}$$

The divergence is given by

$$\begin{aligned}D\left(Geo^{\theta_1}\middle\| Geo^{\theta_2}\right) &= \int_{\theta_1}^{\theta_2}\frac{\mu-\theta_1}{\mu+\mu^2}\,\mathrm{d}\mu \\ &= \theta_1\ln\frac{\theta_1}{\theta_2}-\left(\theta_1+1\right)\ln\frac{\theta_1+1}{\theta_2+1}.\end{aligned}$$

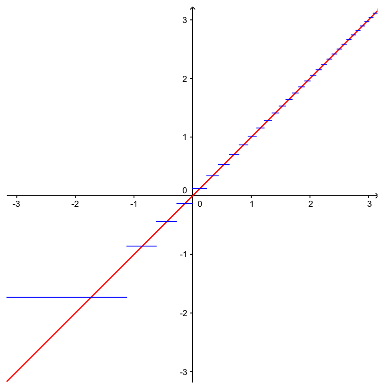

Fig. 5. Plot the quantiles of the signed log-likelihood of $Exp^{3.5}$ vs. the quantiles of the signed log-likelihood of $Geo^{3.5}$.

Hence the signed log-likelihood of the geometric distribution with mean θ is given by

$$g_\theta(x) = \pm\left[2\left(x\ln\frac{x}{\theta} - (x+1)\ln\frac{x+1}{\theta+1}\right)\right]^{1/2}. \tag{1}$$

A QQ-plot of the distributions of the signed log-likelihood of an exponential distribution and a geometric distribution can be seen in Figure 5 and as one can see we get a nice pattern that we will now formalize.

Theorem 5.1. Assume that the random variable M has a geometric distribution Geo^θ and let the random variable X be exponentially distributed Exp^θ. If

$$\Pr(X \le x) = \Pr(M < m)$$

then

$$G_{Geo^\theta}(m-1) \le G_{Exp^\theta}(x) \le G_{Geo^\theta}(m). \tag{2}$$

Proof. First we note that $G_{Exp^\theta}(x) = \gamma(x/\theta)$ and $\Pr(X \le x) = \Pr(X/\theta \le x/\theta)$. Therefore we introduce the variable $y = x/\theta$ and the random variable $Y = X/\theta$ that is exponentially distributed Exp^1.

We will prove that

$$\Pr(Y \le y) = \Pr(M < m) \tag{3}$$

implies

$$g_\theta(m-1) \le \gamma(y) \le g_\theta(m).$$

One has to prove that $\Pr(Y \le y) = \Pr(M < m)$ implies that $g_\theta(m-1) \le \gamma(y)$. Equivalently we have to prove that

$$\gamma(y) - g_\theta(m-1) = \frac{\gamma(y)^2 - g_\theta(m-1)^2}{\gamma(y) + g_\theta(m-1)}$$

is positive. The probability $\Pr(M < m)$ is a decreasing function of θ. Therefore the probability $\Pr(Y \leq y)$ is a decreasing function of θ, but the distribution of Y does not depend on θ so y must be a decreasing function of θ. Therefore the denominator $\gamma(y) + g_\theta(m-1)$ is a decreasing function of θ and it equals zero when $\theta = m-1$. The numerator also equals zero when $\theta = m-1$ so it is sufficient to prove that the numerator is a decreasing function of θ. Therefore we have to prove the inequality

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\gamma(y)^2 - g_\theta(m-1)^2\right) \leq 0$$

or, equivalently, that

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\left(g_\theta(m-1)^2\right) \geq \frac{\mathrm{d}}{\mathrm{d}\theta}\left(\gamma(y)^2\right).$$

One also have to prove that $\Pr(Y \leq y) = \Pr(M < m)$ implies that $\gamma(y) \leq g_\theta(m)$ and it is sufficient to prove that

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\gamma(y)^2\right) \geq \frac{\mathrm{d}}{\mathrm{d}\theta}\left(g_\theta(m)^2\right).$$

We have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\gamma(y)^2\right) &= \frac{\mathrm{d}y}{\mathrm{d}\theta}\cdot\frac{\mathrm{d}}{\mathrm{d}y}\left(\gamma(y)^2\right)\\ &= \frac{\mathrm{d}y}{\mathrm{d}\theta}\cdot 2\left(1-\frac{1}{y}\right).\end{aligned}$$

For the geometric distribution we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\theta}\left((g_\theta(m))^2\right) &= \frac{\mathrm{d}}{\mathrm{d}\theta}\left(2\left(m\ln\frac{m}{\theta} - (m+1)\cdot\ln\frac{m+1}{\theta+1}\right)\right)\\ &= 2\left(-\frac{m}{\theta}+\frac{m+1}{\theta+1}\right)\\ &= 2\frac{\theta-m}{\theta+\theta^2}.\end{aligned}$$

Therefore we have to prove that

$$2\frac{\theta-m+1}{\theta+\theta^2} \geq 2\frac{\mathrm{d}y}{\mathrm{d}\theta}\cdot\left(1-\frac{1}{y}\right) \geq 2\frac{\theta-m}{\theta+\theta^2}.$$

Equation (3) can be solved as

$$\begin{aligned}1-\exp(-y) &= 1-\left(\frac{\theta}{\theta+1}\right)^m\\ y &= m\ln\left(\frac{\theta+1}{\theta}\right).\end{aligned}$$

The derivative is

$$\begin{aligned}\frac{\mathrm{d}y}{\mathrm{d}\theta} &= m\left(\frac{1}{\theta+1}-\frac{1}{\theta}\right)\\ &= -\frac{m}{\theta+\theta^2}.\end{aligned}$$

Finally we have to prove that

$$\begin{aligned}\frac{\theta-m+1}{\theta+\theta^2} &\geq -\frac{m}{\theta+\theta^2}\cdot\left(1-\frac{1}{m\ln\left(\frac{\theta+1}{\theta}\right)}\right)\geq\frac{\theta-m}{\theta+\theta^2}\\ \theta-m+1 &\geq -m+\frac{1}{\ln\left(\frac{\theta+1}{\theta}\right)}\geq\theta-m\\ \theta+1 &\geq \frac{1}{\ln\left(\frac{\theta+1}{\theta}\right)}\geq\theta\\ (\theta+1)\ln\left(\frac{\theta+1}{\theta}\right) &\geq 1\geq\theta\ln\left(1+\frac{1}{\theta}\right),\end{aligned}$$

which is easily checked. □

Corollary 5.2. Assume that the random variable M has a geometric distribution Geo^θ and let the random variable X be exponential distributed Exp^θ. If

$$G_{Exp^\theta}(x)=G_{Geo^\theta}(m)$$

then

$$\Pr(M<m)\leq\Pr(X\leq x)\leq\Pr(M\leq m).$$

If we plot quantiles of an exponential distribution against the corresponding quantiles of the signed log-likelihood of a geometric distribution we get a staircase function, i. e. a sequence of horizontal lines. The inequality means that the left endpoint of any step is to the left of the line $y=x$ and that each right endpoint is to the right of the line. Actually the line $y=x$ intersects each step and we say that the plot has an *intersection property* as illustrated in Figure 5.

Proof. According to Theorem 5.1 we have the implication

$$\Pr(X\leq x)=\Pr(M<m)\text{ implies }G_{Exp^\theta}(x)\leq G_{Geo^\theta}(m).$$

Both $\Pr(X\leq x)$ and $G_{Exp^\theta}(x)$ are increasing functions of x so the previous implication is equivalent to the following implication

$$G_{Exp^\theta}(x)=G_{Geo^\theta}(m)\text{ implies }\Pr(M<m)\leq\Pr(X\leq x).$$

Since

$$\Pr(X\leq x)=\Pr(M<m)\ \text{ implies }G_{Geo^\theta}(m-1)\leq G_{Exp^\theta}(x)$$

we have that $G_{Geo^\theta}(m-1) = G_{Exp^\theta}(x)$ implies that $\Pr(X \leq x) \leq \Pr(M < m)$. Hence $G_{Geo^\theta}(m+1) = G_{Exp^\theta}(x)$ implies that

$$\Pr(X \leq x) \leq \Pr(M < m+1) = \Pr(M \leq m).$$

Since $G_{Geo^\theta}(m) \leq G_{Geo^\theta}(m+1)$ we also have that $G_{Geo^\theta}(m) = G_{Exp^\theta}(x)$ implies that $\Pr(X \leq x) \leq \Pr(M \leq m)$. □

6. INEQUALITIES FOR NEGATIVE BINOMIAL DISTRIBUTIONS

Compounding a Poisson distribution $Po(\lambda)$ with rate parameter λ distributed according to a Gamma distribution $\Gamma(k,\theta)$ leads a *negative binomial distribution.* The link to waiting times in Bernoulli processes will be explored in Section 7. In this section we will parametrize the negative binomial distribution as $neg(k,\theta)$ where k and θ are the parameters of the corresponding Gamma distribution. We note that this is an unusual parametrization the negative binomial distribution, but it will be useful for our calculations. Since λ is both the mean and the variance of $Po(\lambda)$ we can calculate the mean of $neg(k,\theta)$ as $\mu = k\theta$ and the variance as $V(\mu) = \mu + \frac{\mu^2}{k}$.

The point probabilities of a negative binomial distribution can be written in the following way

$$\begin{aligned}\Pr(M=m) &= \int_0^\infty \frac{\lambda^m}{m!}\exp(-\lambda)\cdot\frac{1}{\theta^k}\frac{1}{\Gamma(k)}\lambda^{k-1}\exp\left(-\frac{\lambda}{\theta}\right)\,\mathrm{d}\lambda\\ &= \int_0^\infty \frac{(\theta t)^m}{m!}\exp(-\theta t)\cdot\frac{1}{\Gamma(k)}t^{k-1}\exp(-t)\,\mathrm{d}t\\ &= \frac{\Gamma(m+k)}{m!\Gamma(k)}\cdot\frac{\theta^m}{(\theta+1)^{m+k}}.\end{aligned}$$

The divergence is given by

$$\begin{aligned}D(neg(k,\theta_1)\|\,neg(k,\theta_2)) &= \int_{k\theta_1}^{k\theta_2}\frac{\mu-k\theta_1}{\mu+\frac{\mu^2}{k}}\,\mathrm{d}\mu\\ &= k\left(\theta_1\ln\frac{\theta_1}{\theta_2}-(\theta_1+1)\ln\frac{\theta_1+1}{\theta_2+1}\right).\end{aligned}$$

The signed log-likelihood is given by

$$G_{neg(k,\theta)}(x) = k^{1/2}g_\theta\left(\frac{x}{k}\right)$$

where g_θ is given by Equation (1).

We will need the following lemma.

Lemma 6.1. A Poisson random variable K with distribution $Po(\lambda)$ satisfies

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\Pr(K\leq k) = -\Pr(K=k).$$

Proof. We have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\lambda}\Pr(K\le k) &= \frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\sum_{m=0}^{k}\frac{\lambda^m}{m!}\exp(-\lambda)\right)\\ &= -\exp(-\lambda)+\sum_{m=1}^{k}\left(\frac{\lambda^{(m-1)}}{(m-1)!}\exp(-\lambda)-\frac{\lambda^m}{m!}\exp(-\lambda)\right)\\ &= -\frac{\lambda^k}{k!}\exp(-\lambda),\end{aligned}$$

which proves the lemma. □

Lemma 6.2. If the distribution of M_k is $neg\,(k,\theta)$ then the derivative of the point probability with respect to the mean value parameter equals

$$\frac{\mathrm{d}}{\mathrm{d}\mu}\Pr(M_k\le m) = -\Pr(M_{k+1}=m).$$

where M_{k+1} is $neg\,(k+1,\theta)$.

Proof. We have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\mu}\Pr(M_k\le m) &= \frac{1}{\frac{\mathrm{d}\mu}{\mathrm{d}\theta}}\cdot\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\int_0^\infty\left(\sum_{j=0}^{m}Po(\theta t;j)\right)\cdot\frac{1}{\Gamma(k)}t^{k-1}\exp(-t)\,\mathrm{d}t\right)\\ &= \frac{1}{k}\cdot\int_0^\infty(-t\cdot Po(\theta t;m))\cdot\frac{1}{\Gamma(k)}t^{k-1}\exp(-t)\,\mathrm{d}t\\ &= -\int_0^\infty Po(\theta t;m)\cdot\frac{1}{\Gamma(k+1)}t^{k}\exp(-t)\,\mathrm{d}t.\end{aligned}$$

The last integral equals $-\Pr(M_{k+1}=m)$, which proves the lemma. □

The following theorem generalizes Corollary 5.2 from $k=1$ to arbitrary positive values of k. We cannot use the same proof technique because we do not have an explicit formula for the quantile function for the Gamma distributions except in the case when $k=1$. Lemma 3.2 cannot be used because we want to compare a discrete distribution with a continuous function. Instead the proof combines a proof method developed by Zubkov and Serov [11] with the ideas and results developed in the previous sections.

Theorem 6.3. Assume that the random variable M has a negative binomial distribution $neg\,(k,\theta)$ and let the random variable X be Gamma distributed $\Gamma(k,\theta)$. If

$$G_{\Gamma(k,\theta)}(x) = G_{neg(k,\theta)}(m)$$

then

$$\Pr(M<m)\le\Pr(X\le x)\le\Pr(M\le m). \tag{4}$$

Proof. Below we only give the proof of the upper bound in Inequality 4. The lower bound is proved the in the same way.

First we note that $G_{\Gamma(k,\theta)}(x) = G_{\Gamma\left(k,\frac{1}{k}\right)}(x/(k\theta))$ and

$$\Pr(X \le x) = \Pr\left(\frac{X}{k\theta} \le \frac{x}{k\theta}\right).$$

Therefore we introduce the variable $y = x/(k\theta)$ and the random variable $Y = X/(k\theta)$ that is Gamma distributed $\Gamma(k, {}^{1}\!/\!{}_{k})$. Introduce the difference

$$\delta(\mu_0) = \Pr(M \le m) - \Pr(Y \le y)$$

where μ_0 is the mean value of M. Note that

$$\lim_{\mu_0 \to 0} \delta(\mu_0) = \lim_{\mu_0 \to \infty} \delta(\mu_0) = 0. \tag{5}$$

Note that there exists (at least) one value of μ_0 such that $\frac{\mathrm{d}\delta}{\mathrm{d}\mu_0} = 0$. It is sufficient to prove that δ is first increasing and then decreasing in $[0, \infty[$.

According to Lemma 6.2 the derivative of $\Pr(M \le m)$ with respect to μ_0 is

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\mu_0}\Pr(M \le m) &= -\frac{\Gamma(m+k+1)}{m!\Gamma(k+1)} \cdot \frac{\theta^m}{(\theta+1)^{m+k+1}}. \\
&= -\frac{m+k}{k(\theta+1)} \cdot \frac{\Gamma(m+k)}{m!\Gamma(k)} \frac{\theta^m}{(\theta+1)^{m+k}} \\
&= -\frac{\hat{\theta}+1}{\theta+1} \cdot \Pr(M = m)
\end{aligned}$$

where $\theta = \mu_0/k$ is the scale parameter and where and $\hat{\theta} = {}^{m}\!/\!{}_{k}$ is the maximum likelihood estimate of the scale parameter. Let $\widehat{\Pr}$ denote the probability of M calculated with respect to this maximum likelihood estimate $\hat{\theta}$. Then we have

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\Pr(M \le m) = -\frac{m+k}{\theta+1}\exp(-D)\,\widehat{\Pr}(M = m).$$

The condition

$$G_{\Gamma(k,\theta)}(x) = G_{neg(k,\theta)}(m)$$

can be written as

$$k^{1/2}\gamma(y) = k^{1/2} g_\theta\left(\hat{\theta}\right),$$

which implies

$$(\gamma(y))^2 = \left(g_\theta\left(\hat{\theta}\right)\right)^2.$$

Differentiation with respect to θ gives

$$2\left(1 - \frac{1}{y}\right)\frac{\mathrm{dy}}{\mathrm{d}\theta} = 2\frac{\theta - \hat{\theta}}{\theta + \theta^2}$$

so that

$$\frac{\mathrm{d}y}{\mathrm{d}\theta} = \frac{y}{y-1} \cdot \frac{\theta - \hat{\theta}}{\theta + \theta^2}.$$

Therefore

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\theta} \Pr(Y \le y) &= f(y) \cdot \frac{\mathrm{d}y}{\mathrm{d}\theta} \\
&= \frac{k^k \exp(-k)}{\Gamma(k)\, k^{1/2}} \cdot \frac{\exp(-D)}{y} \cdot \frac{y}{y-1} \cdot \frac{\theta - \hat{\theta}}{\theta + \theta^2} \\
&= \frac{k^k \exp(-k)}{\Gamma(k)\, k^{1/2}} \cdot \frac{\exp(-D)}{\theta y - \theta} \cdot \frac{\theta - \hat{\theta}}{1 + \theta}.
\end{aligned}$$

Combining these results we get

$$\begin{aligned}
\frac{\mathrm{d}\delta}{\mathrm{d}\theta} &= -\frac{m+k}{\theta+1} \widehat{\Pr}(M = m) \cdot \exp(-D) - \frac{k^k \exp(-k)}{\Gamma(k)\, k^{1/2}} \cdot \frac{\exp(-D)}{\theta y - \theta} \cdot \frac{\theta - \hat{\theta}}{1+\theta} \\
&= \frac{k^k \exp(-k)}{\Gamma(k)} \cdot \frac{\exp(-D)}{\theta + 1} \cdot \left(\frac{\hat{\theta} - \theta}{\theta y - \theta} - \frac{\Gamma(k)(m+k)}{k^k \exp(-k)} \cdot \widehat{\Pr}(M = m) \right).
\end{aligned}$$

Remark that the first factor is positive and that the value of

$$\frac{\Gamma(k)(m+k)}{k^k \exp(-k)} \cdot \widehat{\Pr}(M = m)$$

does not depend on θ. Therefore it is sufficient to prove that $\frac{\hat{\theta}-\theta}{\theta y - \theta}$ is a decreasing function of θ.

The derivative with respect to θ is

$$\frac{-(\theta \cdot y - \theta) - \left(\hat{\theta} - \theta\right)\left(y + \theta \cdot \frac{y}{y-1} \cdot \frac{\theta - \hat{\theta}}{\theta + \theta^2} - 1\right)}{(\theta y - \theta)^2} = \frac{\frac{\left(\hat{\theta}-\theta\right)^2}{\hat{\theta}(1+\theta)(y-1)} - \frac{y-1}{y}}{\frac{(\theta y - \theta)^2}{\hat{\theta} \cdot y}}.$$

We have to prove that

$$\frac{y-1}{y} \ge \frac{\left(\hat{\theta} - \theta\right)^2}{\hat{\theta}(1+\theta)(y-1)}.$$

If $\hat{\theta} \ge \theta$ the inequality is equivalent to

$$\frac{(y-1)^2}{y} \ge \frac{\left(\hat{\theta} - \theta\right)^2}{\hat{\theta}(1+\theta)}.$$

If $\hat{\theta} < \theta$ the inequality is equivalent to

$$\frac{(y-1)^2}{y} \le \frac{\left(\hat{\theta} - \theta\right)^2}{\hat{\theta}(1+\theta)}.$$

The equation $\frac{(y-1)^2}{y} = t$ can be solved with respect to y, which gives the solutions $y = 1 + \frac{t}{2} \pm \frac{\left[t^2+4t\right]^{1/2}}{2}$. For $\hat{\theta} \geq \theta$ we get

$$y \geq 1 + \frac{\frac{\left(\hat{\theta}-\theta\right)^2}{\hat{\theta}(1+\theta)}}{2} + \frac{\left[\left(\frac{\left(\hat{\theta}-\theta\right)^2}{\hat{\theta}(1+\theta)}\right)^2 + 4\frac{\left(\hat{\theta}-\theta\right)^2}{\hat{\theta}(1+\theta)}\right]^{1/2}}{2}$$
$$= 1 + \left(\hat{\theta} - \theta\right) \frac{\hat{\theta} - \theta + \left[\left(\hat{\theta} + \theta\right)^2 + 4\hat{\theta}\right]^{1/2}}{2\hat{\theta}\left(1+\theta\right)}.$$

For $\hat{\theta} < \theta$ we get

$$y \geq 1 + \frac{\frac{\left(\hat{\theta}-\theta\right)^2}{\hat{\theta}(1+\theta)}}{2} - \frac{\left[\left(\frac{\left(\hat{\theta}-\theta\right)^2}{\hat{\theta}(1+\theta)}\right)^2 + 4\frac{\left(\hat{\theta}-\theta\right)^2}{\hat{\theta}(1+\theta)}\right]^{1/2}}{2}$$
$$= 1 + \left(\hat{\theta} - \theta\right) \frac{\hat{\theta} - \theta + \left[\left(\hat{\theta} + \theta\right)^2 + 4\hat{\theta}\right]^{1/2}}{2\hat{\theta}\left(1+\theta\right)}.$$

Since γ is increasing and $\gamma(y) = g_\theta\left(\hat{\theta}\right)$ we have to prove that

$$g_\theta\left(\hat{\theta}\right) \geq \gamma\left(1 + \left(\hat{\theta} - \theta\right) \frac{\hat{\theta} - \theta + \left[\left(\hat{\theta} + \theta\right)^2 + 4\hat{\theta}\right]^{1/2}}{2\hat{\theta}\left(1+\theta\right)}\right)$$

or, equivalently, that

$$g_\theta\left(\hat{\theta}\right) - \gamma\left(1 + \left(\hat{\theta} - \theta\right) \frac{\hat{\theta} - \theta + \left[\left(\hat{\theta} + \theta\right)^2 + 4\hat{\theta}\right]^{1/2}}{2\hat{\theta}\left(1+\theta\right)}\right)$$
$$= \frac{\left\{g_\theta\left(\hat{\theta}\right)\right\}^2 - \left\{\gamma\left(1 + \left(\hat{\theta} - \theta\right) \frac{\hat{\theta}-\theta+\left[\left(\hat{\theta}+\theta\right)^2+4\hat{\theta}\right]^{1/2}}{2\hat{\theta}(1+\theta)}\right)\right\}^2}{g_\theta\left(\hat{\theta}\right) + \gamma\left(1 + \left(\hat{\theta} - \theta\right) \frac{\hat{\theta}-\theta+\left[\left(\hat{\theta}+\theta\right)^2+4\hat{\theta}\right]^{1/2}}{2\hat{\theta}(1+\theta)}\right)}$$

is positive. Both the denominator and the numerator are zero when $\theta = \hat{\theta}$. Therefore it is sufficient to prove that both the denominator and the numerator are decreasing functions of θ.

First we prove that the denominator is decreasing. The first term is obviously decreasing. The second term is composed of γ, which is increasing, and $t \curvearrowright 1+\frac{t}{2}\pm\frac{\left[t^2+4t\right]^{1/2}}{2}$ which is increasing or decreasing depending on the sign of $\pm$, and the function $\theta \curvearrowright \frac{\left(\hat{\theta}-\theta\right)^2}{\hat{\theta}(1+\theta)}$ which is decreasing when $\theta \leq \hat{\theta}$ and increasing when $\theta \geq \hat{\theta}$. Therefore the composed function is a decreasing function of θ.

The numerator can be written as

$$2\left\{\hat{\theta}\ln\frac{\hat{\theta}}{\theta}-\left(\hat{\theta}+1\right)\ln\frac{\hat{\theta}+1}{\theta+1}\right\}-2\left\{\begin{array}{c}\left(\hat{\theta}-\theta\right)\frac{\hat{\theta}-\theta+\left[\left(\hat{\theta}+\theta\right)^2+4\hat{\theta}\right]^{1/2}}{2\hat{\theta}(1+\theta)}\\ -\ln\left(1+\left(\hat{\theta}-\theta\right)\frac{\hat{\theta}-\theta+\left[\left(\hat{\theta}+\theta\right)^2+4\hat{\theta}\right]^{1/2}}{2\hat{\theta}(1+\theta)}\right)\end{array}\right\}.$$

We calculate the derivative with respect to θ, which can be written as

$$\frac{-4\frac{2\theta+\hat{\theta}+4}{\theta+\theta^2}\left(\theta-\hat{\theta}\right)^2}{\theta\left(\theta+\hat{\theta}+2\right)\left[\left(\hat{\theta}+\theta\right)^2+4\hat{\theta}\right]^{1/2}+(\theta+2)\left(\left(\hat{\theta}+\theta\right)^2+4\hat{\theta}\right)},$$

which is obviously less than or equal to zero. □

If we want to give lower bounds and upper bounds to the tail probabilities of a negative binomial distribution the following reformulation of Theorem 6.3 is useful.

Corollary 6.4. *Assume that the random variable M has a negative binomial distribution $neg\left(k,\theta\right)$ and let the random variable X be Gamma distributed $\Gamma\left(k,\theta\right)$. Then*

$$\Pr\left(X\leq x_m\right)\leq\Pr\left(M\leq m\right)\leq\Pr\left(X\leq x_{m+1}\right) \tag{6}$$

where x_m and x_{m+1} are determined by

$$G_{\Gamma(k,\theta)}\left(x_m\right)=G_{neg(k,\theta)}\left(m\right),$$
$$G_{\Gamma(k,\theta)}\left(x_{m+1}\right)=G_{neg(k,\theta)}\left(m+1\right).$$

7. INEQUALITIES FOR BINOMIAL DISTRIBUTIONS AND POISSON DISTRIBUTIONS

We will prove that intersection results for binomial distributions and Poisson distributions follow from the corresponding intersection result for negative binomial distributions and Gamma distributions. We note that the point probabilities of a negative binomial distribution can be written as

$$\frac{\Gamma\left(m+k\right)}{m!\Gamma\left(k\right)}\cdot\frac{\theta^m}{\left(\theta+1\right)^{m+k}}=\frac{k^{\bar{m}}}{m!}p^k\left(1-p\right)^m$$

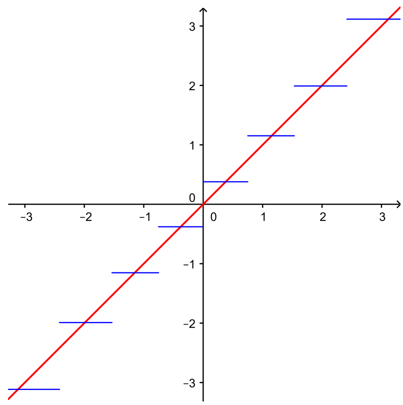

Fig. 6. Plot the quantiles of the signed log-likelihood of a standard Gaussian vs. the quantiles of the signed log-likelihood of $bin\left(7, 1/2\right)$.

where $p = \frac{1}{1+\theta}$ and where $k^{\bar{m}} = k(k+1)(k+2)\dots(k+m-1)$ denotes the raising factorial. Let $nb\left(p,k\right)$ denote a negative binomial distribution with success probability p. Then $nb\left(p,k\right)$ is the distribution of the number of failures before the k'th success in a Bernoulli process with success probability p.

Our inequality for the negative binomial distribution can be translated into an inequality for the binomial distribution. Assume that K is binomial $bin\left(n,p\right)$ and M is negative binomial $nb\left(p,k\right)$. Then

$$\Pr\left(K \geq k\right) = \Pr\left(M + k \leq n\right).$$

In terms of p the divergence can be written as

$$D\left(nb\left(p_1,k\right)\|\, nb\left(p_2,k\right)\right) = \frac{k}{p_1}\left(p_1 \ln\frac{p_1}{p_2} + (1-p_1)\ln\frac{1-p_1}{1-p_2}\right).$$

We have

$$D\left(bin\left(n,p_1\right)\|\, bin\left(n,p_2\right)\right) = n\left(p_1 \ln\frac{p_1}{p_2} + (1-p_1)\ln\frac{1-p_1}{1-p_2}\right)$$

so

$$\begin{aligned} D\left(nb\left(\frac{k}{n},k\right)\middle\|\, nb\left(p,k\right)\right) &= n\left(\frac{k}{n}\ln\frac{\frac{k}{n}}{p_2} + \left(1-\frac{k}{n}\right)\ln\frac{1-\frac{k}{n}}{1-p_2}\right) \\ &= D\left(bin\left(n,\frac{k}{n}\right)\middle\|\, bin\left(n,p\right)\right). \end{aligned}$$

If G_{bin} is the signed log-likelihood of $bin\left(n,p\right)$ and G_{nb} is the signed log-likelihood of $nb\left(p,k\right)$ then $G_{bin(n,p)}\left(k\right) = -G_{nb(p,k)}\left(n-k\right).$

Assume that L is Poisson distributed with mean λ and X is Gamma distributed with shape parameter k and scale parameter 1, i. e. the distribution of the waiting time until k observations from an Poisson process with intensity 1. Then

$$\Pr(L \geq k) = \Pr(X \leq \lambda).$$

Next we note that

$$D(Po(k) \| Po(\lambda)) = D\left(\Gamma\left(k, \frac{\lambda}{k}\right) \middle\| \Gamma(k, 1)\right).$$

If $G_{Po(\lambda)}$ is the signed log-likelihood for $Po(\lambda)$ and $G_{\Gamma(k,1)}$ is the signed log-likelihood for $\Gamma(k, 1)$ then $G_{Po(\lambda)}(k) = -G_{\Gamma(k,1)}(\lambda)$.

Theorem 7.1. Assume that K is binomially distributed $bin(n, p)$ and let $G_{bin(n,p)}$ denote the signed log-likelihood function of the exponential family based on $bin(n, p)$. Assume that L is a Poisson random variable with distribution $Po(\lambda)$ and let $G_{Po(\lambda)}$ denote the signed log-likelihood function of the exponential family based on $Po(\lambda)$. If

$$G_{bin(n,p)}(k) = G_{Po(\lambda)}(k)$$

then

$$\Pr(K < k) \leq \Pr(L < k) \leq \Pr(K \leq k). \tag{7}$$

Proof. Let M denote a negative binomial random variable with distribution $nb(p, k)$ and let X denote a Gamma random variable with distribution $\Gamma(k, \theta)$ where the parameter θ equals $\frac{1}{p} - 1$ such that the distributions $nb(p, k)$ and $\Gamma(k, \theta)$ have the same mean value. Now $G_{nb(p,k)}(n-k) = -G_{bin(n,p)}(k)$ and $G_{\Gamma(k,\theta)}(\lambda\theta) = -G_{Po(\lambda)}(k)$. Then $G_{nb(p,k)}(n-k) = G_{\Gamma(k,\theta)}(\lambda\theta)$. The left part of the Inequality 7 is proved as follows.

$$\begin{aligned}
\Pr(K < k) &= 1 - \Pr(K \geq k) \\
&= 1 - \Pr(M + k \leq n) \\
&\leq 1 - \Pr(X \leq \lambda\theta) \\
&= 1 - \Pr(L \geq k) \\
&= \Pr(L < k).
\end{aligned}$$

The right part of the inequality is proved in the same way. □

Note that Theorem 6.3 cannot be proved from Theorem 7.1 because the number parameter for a binomial distribution has to be an integer while the number parameter of a negative binomial distribution may assume any positive value. Now, our inequalities for negative binomial distributions can be translated into inequalities for binomial distributions.

We can use the previous theorem to give a new proof of an *intersection inequalities* for the binomial family as stated in the following theorem that was recently proved by Zubkov and Serov [11].

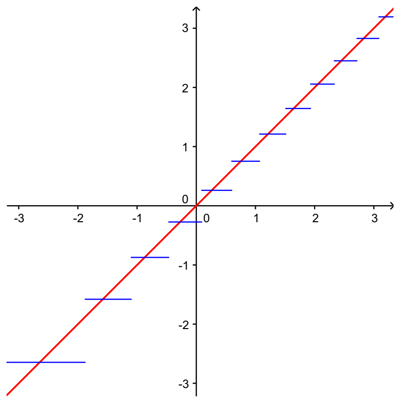

Fig. 7. Plot of quantiles of a standard Gaussian vs. the signed log-ligelihood of the Poisson distribution $Po\,(3.5)$.

Corollary 7.2. Assume that K is binomially distributed $bin\,(n,p)$ and let $G_{bin(n,p)}$ denote the signed log-likelihood function of the exponential family based on $bin\,(n,p)$. Then

$$\Pr\left(K < k\right) \leq \Phi\left(G_{bin(n,p)}\left(k\right)\right) \leq \Pr\left(K \leq k\right). \tag{8}$$

Similarly, assume that L is Poisson distributed $Po\,(\lambda)$ and let $G_{Po(\lambda)}$ denote the signed log-likelihood function of the exponential family based on $Po\,(\lambda)$. Then

$$\Pr\left(L < k\right) \leq \Phi\left(G_{Po(\lambda)}\left(k\right)\right) \leq \Pr\left(L \leq k\right). \tag{9}$$

P r o o f. First we prove the left part of Inequality (9). Let X denote a Gamma distributed $\Gamma\,(k,1)$ and let Z denote a standard Gaussian. Then $G_{Po(\lambda)}\,(k) = -G_{\Gamma(k,1)}\,(\lambda)$ and

$$\begin{aligned}
\Pr\left(L < k\right) &= 1 - \Pr\left(L \geq k\right) \\
&= 1 - \Pr\left(X \leq \lambda\right) \\
&= \Pr\left(X \geq \lambda\right) \\
&\leq \Pr\left(Z \geq G_{\Gamma(k,1)}\left(\lambda\right)\right) \\
&= \Pr\left(Z \geq -G_{Po(\lambda)}\left(k\right)\right) \\
&= \Phi\left(G_{Po(\lambda)}\left(k\right)\right).
\end{aligned}$$

The left part of Inequality (8) is obtained by combining the left part of Inequality (9) with the left part of Inequality (7). The right part of Inequality (8) follows from the left part of Inequality (8) by replacing p by $1-p$ and replacing k by $n-k$. Since a Poisson distribution is a limit of binomial distributions the right part of Inequality (9) follows from the right part of Inequality (9). $\square$

The intersection property of the Poisson distribution can also be proved from the intersection property of the negative binomial distribution and the Gamma distribution by using that a Poisson distribution is a limit of negative binomial distributions and that corresponding Gamma distributions have a Gaussian distribution as limit. The intersection property for Poisson distributions was first proved in [7].

8. SUMMARY

The main theorems in this paper are domination theorems and intersection theorems. Inequalities of the first type state that the signed log-likelihood of one distribution is dominated by the signed log-likelihood of another distribution, i.e. the distribution function of the first distribution is greater than the distribution function of the second distribution.

sgn. log-likelihood	dom. by sgn. log-likelihood	Condition	Theorem
Exponential	Gaussian		3.3
Gamma	Gaussian		4.2
Γ_{k_1,θ_1}	Γ_{k_2,θ_2}	$k_1 \leq k_2$	4.3
Inverse Gaussian	Gaussian		Ref. [6, Thm. 10]
Inv. Gauss(μ_1, λ_1)	Inv. Gauss(μ_2, λ_2)	$\frac{\mu_1}{\lambda_1} > \frac{\mu_2}{\lambda_2}$	Unpublished

Tab. 1. Stochastic domination results. Note that the exponential distributions are special cases of Gamma distributions.

The second type of result are intersection results, i.e. the distribution function of the log-likelihood of a discrete distribution is a staircase function where each step is intersected by the distribution function of the log-likelihood of a continuous distribution.

Discrete distribution	Continuous distribution	Theorem
Geometric	Exponential	5.2
Negative binomial	Gamma	6.3
Binomial	Gaussian	7.2
Poisson	Gaussian	7.2

Tab. 2. Intersection results.

9. DISCUSSION

We have proved that a plot of the quantiles of the signed log-likelihood of an exponential distribution and a geometric distribution satisfies the intersection property via Inequlity (2). With a minor modification of the proof we get the following bound that is much sharper.

$$G_{Geo^\theta}\left(m - {}^1\!/\!_2\right) \leq G_{Exp^\theta}\left(x\right)$$

We conjecture that a similar inequality holds for any Gamma distribution compared with the corresponding negative binomial distribution.

We have both lower bounds and upper bounds on the Poisson distributions. The upper bound for the Poisson distribution corresponds to the lower bound for the Gamma distribution presented in Theorem 4.2, but the lower bound for the Poisson distribution is translated into a new upper bound for the distribution function of the Gamma distribution. Numerical calculations also indicates that the right hand inequality in Inequality (9) can be improved to

$$\Phi\left(G_{Po(\lambda)}\left(k+\frac{1}{2}\right)\right) \leq \Pr\left(L \leq k\right).$$

This inequality is much tighter than the inequality in (9). Similarly, J. Reiczigel, L. Rejtő and G. Tusnády conjectured that both the lower bound and the upper bound in Inequality 8 can be improved significantly when for $p = 1/2$ [10], and their conjecture has been a major motivation for initializing this research.

For the most important distributions like the binomial distributions, the Poisson distributions, the negative binomial distributions, the inverse Gaussian distributions and the Gamma distributions we can formulate sharp inequalities that hold for any sample size. All these distributions have variance functions that are polynomials of order 2 and 3. Natural exponential families with polynomial variance functions of order at most 3 have been classified [8, 9] and there is a chance that one can formulate and prove sharp inequalities for each of these exponential families. Although there may exist very nice results for the rest of the exponential families with simple variance functions the rest of these exponential families have much fewer applications than the exponential families that have been the subject of the present paper.

In the present paper inequalities have been developed for specific exponential families, but one may hope that a more general inequality may be developed where a bound on the tail is derived directly from the properties of the variance function.

ACKNOWLEDGEMENT

The author want to thank Gabor Tusnády, who inspired me to look at this type of inequalities. The author also want the thank Joram Gat, László Györfi, János Komlós, Sebastian Ziesche, and A. Zubkov for useful correspondence or discussions. Finally I want to thank Narayana Prasad Santhanam for inviting me a two month visit at the Electrical Engineering Department, University of Hawai'i at Mānoa where this paper was completed.

(Received February 23, 2016)

REFERENCES

[1] D. Alfers and H. Dinges: A normal approximation for beta and gamma tail probabilities. Z. Wahrscheinlichkeitstheory verw. Geb. *65* (1984), 3, 399–420. DOI:10.1007/bf00533744

[2] R. R. Bahadur: Some approximations to the binomial distribution function. Ann. Math. Statist. *31* (1960), 43–54. DOI:10.1214/aoms/1177705986

[3] R. R. Bahadur and R. R. Rao: On deviation of the sample mean. Ann. Math. Statist. *31* (1960), 4, 1015–1027. DOI:10.1214/aoms/1177705674

[4] O. E. Barndorff-Nielsen: A note on the standardized signed log likelihood ratio. Scand. J. Statist. *17* (1990), 2, 157–160.

[5] L. Györfi, P. Harremoës, and G. Tusnády: Gaussian approximation of large deviation probabilities. http://www.harremoes.dk/Peter/ITWGauss.pdf, 2012.

[6] P. Harremoës: Mutual information of contingency tables and related inequalities. In: Proc. ISIT 2014, IEEE 2014, pp. 2474–2478. DOI:10.1109/isit.2014.6875279

[7] P. Harremoës and G. Tusnády: Information divergence is more χ^2-distributed than the χ^2-statistic. In: International Symposium on Information Theory (ISIT 2012) (Cambridge, Massachusetts), IEEE 2012, pp. 538–543. DOI:10.1109/isit.2012.6284247

[8] G. Letac and M. Mora: Natural real exponential families with cubic variance functions. Ann. Stat. *18* (1990), 1, 1–37. DOI:10.1214/aos/1176347491

[9] C. Morris: Natural exponential families with quadratic variance functions. Ann. Statist. *10* (1982), 65–80. DOI:10.1214/aos/1176345690

[10] J. Reiczigel, L. Rejtő, and G. Tusnády: A sharpning of Tusnády's inequality. arXiv: 1110.3627v2, 2011.

[11] A. M. Zubkov and A. A. Serov: A complete proof of universal inequalities for the distribution function of the binomial law. Theory Probab. Appl. *57* (2013), 3, 539–544. DOI:10.1137/s0040585x97986138

Peter Harremoës, Copenhagen Business College, Copenhagen. Denmark.
e-mail: harremoes@ieee.org