Bernoulli Distribution and Bernoulli trials

We say X follows a Bernoulli distribution with parameter p if X take only 0 and 1 and the probabilities are $P(X = 1) = p$ and $P(X = 0) = 1 - p$. We write $X \sim Ber(p)$. We have studied the basic properties of $Ber(p)$. For example, $EX = p$, $Var(X) = p(1 - p)$ and its MGF is $\psi(t) = 1 - p + pe^t$.

n **Bernoulli trials:** n independent $Ber(p)$ random variables form n Bernoulli trails.

For each Bernoulli trial, if $X = 1$ then we say we have a success, otherwise we have a failure. So $p = P(X = 1)$ is also called the probability of success.

Binomial Distribution

Definition: We say X follows a Binomial distribution with parameters n and p if

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \; k = 0, 1, \ldots, n$$

We write $X \sim Bin(n, p)$.

We have studied the basic properties of $Bin(n, p)$. For example, $EX = np$, $Var(X) = np(1-p)$ and the MGF is $\psi(t) = (1 - p + pe^t)^n$.

The model Suppose we have n Bernoulli trails with probability of success p. Let Y be the number of success in the n Bernoulli trails. Then $Y \sim Bin(n, p)$.

Why? let X_i be the ith $Ber(p)$ rv in the n Bernoulli trails. Then we observe

$$Y = X_1 + X_2 + \cdots + X_n$$

Geometric Distribution

Definition: We say X follows a Geometric distribution with parameter of success p if

$$P(X = k) = (1-p)^k p, \;\; k = 0, 1, 2, \ldots,$$

We write $X \sim Geo(p)$. For convenience we also let $q = 1 - p$. $P(X = k) = q^k p$.

The model Suppose we keep doing the Bernoulli trails until we have a success. Let X be the number of trials before the success. Then $X \sim Geo(p)$.

MGF of $Geo(p)$

$$\psi(t) = E[e^{tX}] = \sum_{k=0}^{\infty} e^{tk} q^k p = p \sum_{k=0}^{\infty} (e^t q)^k$$

$$\psi(t) = \frac{p}{1 - e^t q} \qquad for \;\; t < \log(1/q)$$

Mean and Variance of $Geo(p)$

If $X \sim Geo(p)$ then

$$EX = \frac{q}{p} \quad Var(X) = \frac{q}{p^2}.$$

Explanation:

$$\psi^{(1)}(t) = \frac{p}{(1-e^t q)^2} q e^t$$

$$EX = \psi^{(1)}(0) = \frac{p}{(1-q)^2} q = \frac{q}{p}$$

$$\psi^{(2)}(t) = \frac{p}{(1-e^t q)^2} q e^t + 2\frac{p}{(1-e^t q)^3} q e^t \cdot q e^t$$

$$E(X^2) = \psi^{(2)}(0) = \frac{p}{(1-q)^2} q + 2\frac{p}{(1-q)^3} q^2$$

$$E(X^2) = \frac{q}{p} + \frac{2q^2}{p^2}$$

$$Var(X) = E(X^2) - (EX)^2 = \frac{q}{p^2}$$

Poisson Distribution

Definition: We say X follows a Poisson distribution with parameter λ if

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!},\ k = 0, 1, 2, \ldots,$$

We write $X \sim Poi(\lambda)$.

MGF of $Poi(\lambda)$: Let $X \sim Poi(\lambda)$ then its MGF is $\psi(t) = e^{\lambda(e^t - 1)}$.

$$\begin{aligned}
\psi(t) &= E(e^t X) = \sum_{k=0}^{\infty} e^{tk}\frac{e^{-\lambda}\lambda^k}{k!} \\
&= e^{-\lambda}\sum_{k=0}^{\infty}\frac{(e^t\lambda)^k}{k!}\ (*) \\
&= e^{-\lambda}e^{e^t\lambda} = e^{\lambda(e^t-1)}.
\end{aligned}$$

Note that we have used the following series

$$e^a = \sum_{k=0}^{\infty}\frac{a^k}{k!}$$

at the step (*) we let $a = e^t\lambda$.

Mean and Variance of $Poi(\lambda)$: Let $X \sim Poi(\lambda)$ then

$$EX = \lambda \quad Var(X) = \lambda.$$

Because

$$\psi^{(1)}(t) = e^{\lambda(e^t-1)}\lambda e^t = \lambda e^{\lambda(e^t-1)+t} \quad \psi^{(1)}(0) = \lambda$$

$$\psi^{(2)}(t) = \lambda e^{\lambda(e^t-1)+t}(\lambda e^t+1) \quad \psi^{(2)}(0) = \lambda(\lambda+1)$$

Thus

$$EX = \psi^{(1)}(0) = \lambda$$

$$Var(X) = \psi^{(2)}(0) - (\psi^{(1)}(0))^2 = \lambda$$

An interesting property: If X_i, $1 \leq i \leq n$ are independent Poisson rvs and $X_i \sim Poi(\lambda_i)$ for $1 \leq i \leq n$. Then $Y = X_1 + X_2 + \ldots + X_n \sim Poi(\sum_{i=1}^n \lambda_i)$. Why?

$$\psi_Y(t) = \prod_{i=1}^n \psi_{X_i}(t) = \prod_{i=1}^n e^{\lambda_i(e^t-1)} = e^{(\sum_{i=1}^n \lambda_i)(e^t-1)}$$

Which is the MGF of $Poi(\sum_{i=1}^n \lambda_i)$. So $Y \sim Poi(\sum_{i=1}^n \lambda_i)$.

Normal Distribution

Definition: We write $X \sim N(\mu, \sigma^2)$ if the pdf of X is

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

μ and σ^2 are the parameters of the normal distribution.

A basic property: If $X \sim N(\mu, \sigma^2)$, let $Z = \frac{X-\mu}{\sigma}$. Then $Z \sim N(0,1)$. Conversely, if $Z \sim N(0,1)$, $X = \sigma Z + \mu \sim N(\mu, \sigma^2)$.

Explanation: $Z = g(X) = \frac{X-\mu}{\sigma}$. $g(x) = \frac{x-\mu}{\sigma}$, its inverse function is $x = g^{-1}(z) = \sigma z + \mu$.

$$\begin{aligned} f_Z(z) &= f_X(g^{-1}(z))|[g^{-1}(z)]'| \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\sigma z + \mu - \mu)^2}{2\sigma^2}} \cdot \sigma \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \end{aligned}$$

Three immediate applications:

1. Mean and variance of $N(\mu, \sigma^2)$:

$E[X] = E[\mu + \sigma Z] = \mu + \sigma E[Z] = \mu + \sigma 0 = \mu.$

$Var[X] = Var[\mu + \sigma Z] = \sigma^2 Var[Z] = \sigma^2 1 = \sigma^2.$

2. MGF of $N(\mu, \sigma^2)$: By the basic property, $X = \sigma Z + \mu$ with $Z \sim N(0, 1)$. Thus

$$\psi_X(t) = e^{t\mu + \frac{\sigma^2}{2} t^2}$$

3. CDF of $X \sim N(\mu, \sigma^2)$:

$$\begin{aligned} P(X \leq x) &= P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) \\ &= P\left(N(0, 1) \leq \frac{x - \mu}{\sigma}\right) \\ &= \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} ds. \end{aligned}$$

Linear combinations of independent normals

Theorem 1: Suppose $X_1, \ldots, X_n$ are n independent normal random variables. Let $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, \ldots, n$. The distribution of $Y = \sum_{i=1}^n a_i X_i + b$, where a_i's are some constants, is a normal with

$$mean = \sum_{i=1}^n a_i \mu_i + b \ , \quad variance = \sum_{i=1}^n a_i^2 \sigma_i^2.$$

Typical examples: X_1 and X_2 are independent normals, and $Y = X_1 - X_2$ or $Y = X_1 + X_2$.

Explanation: We compute the MGF of Y.

$$\begin{aligned}
\psi_Y(t) &= E[e^{t(a_1X_1+\cdots+a_nX_n+b)}] \\
&= e^{tb}E[e^{a_1tX_1}\cdots e^{a_ntX_n}] \\
&= e^{tb}E[e^{a_1tX_1}]\cdots E[e^{a_ntX_n}] \text{ [by independence]} \\
&= e^{tb}\psi_{X_1}(a_1t)\cdots\psi_{X_n}(a_nt) \text{ [by definition]}
\end{aligned}$$

Linear combinations of independent normals, contd'

$$\psi_{X_i}(a_i t) = e^{\mu_i a_i t + \frac{1}{2}\sigma_i^2 (a_i t)^2}$$

$$\begin{aligned}
\psi_Y(t) &= e^{tb}\psi_{X_1}(a_1 t)\cdots\psi_{X_n}(a_n t) \\
&= e^{tb}\prod_{i=1}^{n} e^{\mu_i a_i t + \frac{1}{2}\sigma_i^2 (a_i t)^2} \\
&= e^{tb} e^{\sum_{i=1}^{n} \mu_i a_i t + \frac{1}{2}\sigma_i^2 (a_i t)^2} \\
&= e^{[b+\sum_{i=1}^{n} \mu_i a_i] t + \frac{1}{2}[\sum_{i=1}^{n} \sigma_i^2 a_i^2] t^2}
\end{aligned}$$

Write

$$\mu_Y = \sum_{i=1}^{n} \mu_i a_i + b \quad \sigma_Y^2 = \sum_{i=1}^{n} \sigma_i^2 a_i^2.$$

then $\psi_Y(t) = e^{\mu_Y t + \frac{1}{2}\sigma_Y^2 t^2}$.

The above is the MGF of $N(\mu_Y, \sigma_Y^2)$. Therefore we know $Y \sim N(\mu_Y, \sigma_Y^2)$.

Examples of theorem 1

Suppose $X \sim N(4, 1)$, $Y \sim N(5, 4)$ and $Z \sim N(2, 2)$. They are independent. Find the distribution of the following (a). $X + Y + Z$, (b). $X - Y$, (c) $2X - Y - Z$.

Solution: First, we know that the distributions are normal, because (a)–(c) are linear combination of independent normals. We only need to find out the mean and variance.

(a) $\mu = 4 + 5 + 2 = 11$, $\sigma^2 = 1 + 4 + 2 = 7$.
$X + Y + Z \sim N(11, 7)$.

(b) $\mu = 4 - 5 = -1$, $\sigma^2 = 1 + 4 = 5$.
$X - Y \sim N(-1, 5)$.

(c) $\mu = 2*4-5-1 = 1$, $\sigma^2 = 2^2*1+4+2 = 10$.
$2X - Y - Z \sim N(1, 10)$.

Bivariate Normal Distribution

We say a pair of random variables (X, Y) follow a bivariate normal distribution with parameters $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, \rho)$, if their joint pdf is given by

$$f(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}\sigma_1\sigma_2} \exp\{-\frac{1}{2(1-\rho^2)}$$

$$[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho(\frac{x-\mu_1}{\sigma_1})(\frac{y-\mu_2}{\sigma_2}) + \frac{(y-\mu_2)^2}{\sigma_2^2}]\}$$

where $-1 < \rho < 1$.

When $\rho = 0$, then the joint pdf is

$$\begin{aligned} f(x,y) &= \frac{1}{2\pi\sigma_1\sigma_2} \exp\{-\frac{1}{2}[\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2}]\} \\ &= \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}} \end{aligned}$$

which means that X, Y are independent normal variables and $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$.

Construct Bivariate Normal from Two Independent Normals

Let Z_1 and Z_2 be two independent standard Normal random variables. Then the joint pdf of (Z_1, Z_2) is $f(z_1, z_2) = \frac{1}{2\pi}\exp[-\frac{z_1^2+z_2^2}{2}]$. Now construct two new random variables (X, Y) by the following linear transformation

$$X = \sigma_1 Z_1 + \mu_1 \quad Y = \sigma_2[\rho Z_1 + \sqrt{1-\rho^2} Z_2] + \mu_2.$$

Let's find the joint pdf of (X, Y). Note that the inverse transformation is

$$Z_1 = (X - \mu_1)/\sigma_1 \quad Z_2 = [\frac{Y-\mu_2}{\sigma_2} - \rho\frac{X-\mu_1}{\sigma_1}]/\sqrt{1-\rho^2}$$

The Jacobian is $J = \frac{1}{\sqrt{1-\rho^2}\sigma_1\sigma_2}$. Thus, we can find the joint pdf of (X, Y) is exactly the joint pdf of a bivariate normal with parameters $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, \rho)$. [Do the calculations by yourself!]

Linear Combinations of Bivariate Normal

Let (X, Y) follow bivaraite normal distribution with parameters $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, \rho)$. Consider $Z = aX + bY + c$. Then $Z \sim N(\mu_z, \sigma_z^2)$ where

$$\mu_z = a\mu_1 + b\mu_2 + c\,; \quad \sigma_z^2 = a^2\sigma_1^2 + b^2\sigma_2^2 + 2ab\rho\sigma_1\sigma_2.$$

Explanation: Consider (X^*, Y^*) by

$$X^* = \sigma_1 Z_1 + \mu_1 \quad Y^* = \sigma_2[\rho Z_1 + \sqrt{1-\rho^2} Z_2] + \mu_2.$$

where Z_1, Z_2 are iid $N(0,1)$. Let $Z^* = aX^* + bY^* + c$. Since (X^*, Y^*) follows the same bivariate normal distribution of (X, Y) we know Z^* and Z have the same distribution. We can write Z^* as

$$Z^* = a(\sigma_1 Z_1 + \mu_1) + b(\sigma_2\rho Z_1 + \sigma_2\sqrt{1-\rho^2} Z_2 + \mu_2) + c$$

$$Z^* = (a\sigma_1 + b\sigma_2\rho)Z_1 + (b\sigma_2\sqrt{1-\rho^2})Z_2 + (a\mu_1 + b\mu_2 + c)$$

By Theorem 1 on page 12, $Z^* \sim N(\mu_z, \sigma_z^2)$.

Marginal Distributions and Correlation

Let (X, Y) follow bivaraite normal distribution with parameters $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, \rho)$. Then $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$. Moreover, the correlation between X and Y is ρ.

Explanation: Consider $X = 1 \cdot X + 0 \cdot Y + 0$ and $Y = 0 \cdot X + 1 \cdot Y + 0$. Then use the linear combination result.

$$
\begin{aligned}
& Cov(X, Y) = Cov(X^*, Y^*) \\
= \;& Cov(\sigma_1 Z_1 + \mu_1, \sigma_2[\rho Z_1 + \sqrt{1-\rho^2} Z_2] + \mu_2) \\
= \;& \sigma_1 \sigma_2 \rho . (check\ this\ by\ yourself).
\end{aligned}
$$

$$
Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{\sigma_1^2 \sigma_2^2}} = \frac{\sigma_1 \sigma_2 \rho}{\sigma_1 \sigma_2} = \rho .
$$

Conditional Distributions

Let (X, Y) follow bivaraite normal distribution with parameters $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, \rho)$. Then the conditional distribution of Y given $X = x$ is normal with mean $\mu_2 + \rho\sigma_2(\frac{x-\mu_1}{\sigma_1})$ and variance $(1-\rho^2)\sigma_2^2$.

Explanation: The conditional pdf is computed by $f(y|X = x) = \frac{f_{(X,Y)}(x,y)}{f_X(x)}$. We know $f_X(x) = \frac{e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}}{\sqrt{2\pi\sigma_1^2}}$ because the marginal distribution of X is $N(\mu_1, \sigma_1^2)$. Simplifying the ratio we can find that $f(y|X = x)$ is the pdf of normal with mean $\mu_2 + \rho\sigma_2(\frac{x-\mu_1}{\sigma_1})$ and variance $(1-\rho^2)\sigma_2^2$.

Question: what is the conditional pdf of X given $Y = y$? Answer: It is normal with mean $\mu_1 + \rho\sigma_1(\frac{y-\mu_2}{\sigma_2})$ and variance $(1-\rho^2)\sigma_1^2$.

Gamma Distribution

Definition We say X follows a Gamma distribution with parameters (α, λ), if the pdf of X is

$$f(x|\alpha,\lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \ x > 0$$

where $\Gamma(\alpha)$ is the Gamma function defined by

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx.$$

Write $X \sim Gam(\alpha, \lambda)$. $\alpha > 0$ is the shape parameter, $\lambda > 0$ is the scale parameter.

MGF $Gam(\alpha, \lambda)$

Suppose $X \sim Gam(\alpha, \lambda)$, then

$$\psi_X(t) = \left(\frac{1}{1 - \frac{t}{\lambda}}\right)^\alpha \quad t < \lambda$$

$$\begin{aligned}
\psi_X(t) &= \int_0^\infty e^{tx} x^{\alpha-1} e^{-\lambda x} dx \\
&= \lambda^\alpha \int_0^\infty \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-(\lambda-t)x} dx \\
&= \frac{\lambda^\alpha}{(\lambda-t)^\alpha} [\int_0^\infty \frac{(\lambda-t)^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-(\lambda-t)x} dx] \\
&= \frac{\lambda^\alpha}{(\lambda-t)^\alpha} \cdot 1 \\
&= \left(\frac{1}{1 - \frac{t}{\lambda}}\right)^\alpha
\end{aligned}$$

Note that inside [] is the integral of the pdf of $Gam(\alpha, \lambda - t)$

Mean and Variance of $Gam(\alpha, \lambda)$

Suppose $X \sim Gam(\alpha, \lambda)$, then

$$E(X) = \frac{\alpha}{\lambda} \quad Var(X) = \frac{\alpha}{\lambda^2}$$

Explanation:

$$\psi^{(1)}(t) = (-\alpha)(1-\frac{t}{\lambda})^{-\alpha-1}(-\frac{1}{\lambda}) = \frac{\alpha}{\lambda}(1-\frac{t}{\lambda})^{-\alpha-1}$$

Thus

$$EX = \psi^{(1)}(0) = \frac{\alpha}{\lambda}.$$

Similarly,

$$\psi^{(2)}(t) = \frac{\alpha}{\lambda}\frac{\alpha+1}{\lambda}(1-\frac{t}{\lambda})^{-\alpha-2}$$

thus

$$E(X^2) = \psi^{(2)}(0) = \frac{\alpha}{\lambda}\frac{\alpha+1}{\lambda}.$$

Then

$$Var(X) = E(X^2)-(EX)^2 = \frac{\alpha}{\lambda}\frac{\alpha+1}{\lambda}-(\frac{\alpha}{\lambda})^2 = \frac{\alpha}{\lambda^2}.$$

Sum of independent Gamma RVs

Theorem 4: Suppose $X_1, \ldots, X_n$ are n independent $Gamma$ random variables. Let $X_i \sim Gam(\alpha_i, \lambda)$, $i = 1, \ldots, n$. The distribution of $Y = \sum_{i=1}^n X_i$ is $Gam(\sum_{i=1}^n \alpha_i, \lambda)$.

Explanation: check the MGF of Y.

$$\psi_Y(t) = \prod_{i=1}^n \Phi_{X_i}(t) = \prod_{i=1}^n \left(\frac{1}{1-\frac{t}{\lambda}}\right)^{\alpha_i}$$

$$\psi_Y(t) = \left(\frac{1}{1-\frac{t}{\lambda}}\right)^{\sum_{i=1}^n \alpha_i}$$

which is the MGF of $Gam(\sum_{i=1}^n \alpha_i, \lambda)$. Thus $Y = \sum_{i=1}^n X_i$ is $Gam(\sum_{i=1}^n \alpha_i, \lambda)$.

Exponential Distribution

Definition: We say X follows an exponential distribution with parameter λ, if $X \sim Gam(1, \lambda)$. We write $X \sim Exp(\lambda)$.

$$f(x) = \lambda e^{-\lambda x}, x > 0$$

When $\lambda = 1$, the distribution is called the standard exponential distribution.

Mean, Variance and MGF of $Exp(\lambda)$

If $X \sim Exp(\lambda)$, then we know $X \sim Gam(1, \lambda)$, thus

$$EX = \frac{\alpha}{\lambda} = \frac{1}{\lambda} \quad Var(X) = \frac{\alpha}{\lambda^2} = \frac{1}{\lambda^2}$$

The MGF of X is

$$\psi(t) = (\frac{1}{1 - \frac{t}{\lambda}})^{\alpha} = \frac{1}{1 - \frac{t}{\lambda}}$$

Chi-square Distribution

Definition: We say X follows a Chi-square distribution with degrees of freedom k, if $X \sim Gam(\frac{k}{2}, \frac{1}{2})$. We write $X \sim \chi_k^2$.

$$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}, \ x > 0.$$

Mean, Variance and MGF of χ_k^2

Suppose $X \sim \chi_k^2$ then we know $X \sim Gam(\frac{k}{2}, \frac{1}{2})$. Thus

$$E(X) = \frac{\alpha}{\lambda} = \frac{k/2}{1/2} = k$$

$$Var(X) = \frac{\alpha}{\lambda^2} = \frac{k/2}{1/4} = 2k.$$

$$\psi_X(t) = \left(\frac{1}{1 - \frac{t}{1/2}}\right)^{k/2} = \left(\frac{1}{\sqrt{1-2t}}\right)^k.$$

Sum of independent Chi-squares

Theorem 5: Suppose $X_i \sim \chi^2_{k_i}$ $i = 1, 2, \ldots, n$ are independent Chi-square random variables. Let

$$Y = X_1 + X_2 + \cdots + X_n.$$

Then $Y \sim \chi^2_{\sum_{i=1}^n k_i}$.

Explanation: We write $X_i \sim \Gamma(k_i/2, 1/2)$. Then Y is the sum of independent Gamma RVs with a common scale parameter $\lambda = 0.5$. So Theorem 4 tells us that

$$Y \sim \Gamma(\sum_{i=1}^{n} k_i, 1/2).$$

Then by the definition of Chi-square we know $Y \sim \chi^2_{\sum_{i=1}^n k_i}$.

Distribution	Mean	Variance	MGF
$Ber(p)$	p	$p(1-p)$	$1-p+pe^t$
$Bin(n,p)$	np	$np(1-p)$	$(1-p+pe^t)^n$
$Poi(\lambda)$	λ	λ	$e^{\lambda(e^t-1)}$
$N(\mu,\sigma^2)$	μ	σ^2	$e^{\mu t+\frac{t^2\sigma^2}{2}}$
$Gam(\alpha,\lambda)$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$	$\left(\frac{1}{1-\frac{t}{\lambda}}\right)^\alpha$
$Exp(\lambda)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{1}{1-\frac{t}{\lambda}}$
χ_k^2	k	$2k$	$\left(\frac{1}{\sqrt{1-2t}}\right)^k$

Negative Binomial Distribution

Definition: Let $X_1, \ldots, X_r$ be r iid $Geo(p)$ random variables. Then $Y = \sum_{i=1}^r X_i$ is called a negative binomial random variable and its distribution is called the negative binomial distribution.

By definition,

$$EY = \sum_{i=1}^{r} EX_i = rEX_1 = r\frac{q}{p}$$

By independence, we have

$$Var(Y) = \sum_{i=1}^{r} Var(X_i) = rVar(X_1) = r\frac{q}{p^2}$$

$$\psi_Y(t) = \prod_{i=1}^{r} \psi_{X_i}(t) = [\psi_{X_1}(t)]^r = \left(\frac{p}{1 - e^t q}\right)^r.$$

Lognormal Distribution

Definition If log(Y) follows a normal distribution with mean μ and variance σ^2, Then we say Y follows a lognormal distribution with parameters (μ, σ^2). In other words, if $X \sim N(\mu, \sigma^2)$ then e^X follows a lognormal distribution with parameters (μ, σ^2).

Let Y be a lognormal rv with parameters (μ, σ^2). Then $X = \log(Y) \sim N(\mu, \sigma^2)$.

$$E[Y] = E[e^X] = \psi_X(1) = e^{\mu + \frac{\sigma^2}{2}}$$

$$E[Y^2] = E[e^{2X}] = \psi_X(2) = e^{2\mu + 2\sigma^2}$$

$$Var(Y) = E[Y^2] - (EY)^2 = e^{2\mu + \sigma^2}[e^{\sigma^2} - 1].$$

Beta Distribution

Definition We say X has a beta distribution with parameters α and β ($\alpha > 0, \beta > 0$) if the pdf of X is

$$f(x|\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1},\ 0 < x < 1.$$

Beta distributions are often used to model rvs that take values between 0 and 1.

Mean and Variance Beta Distributions

If X has a beta distribution with parameters α and β, then

$$EX = \frac{\alpha}{\alpha + \beta}$$

$$Var(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

The computations use the properties of $\Gamma(\cdot)$ functions. If you are interested, read page 305-306 and Theorem 5.9.2 on page 296.