Probability and Statistics

Fall 2000

Lect. 6: Special Distributions II (CB 3.2–3.3)

1. Gamma Distribution Using the same metaphor of observing a system with random events occurring, where the exponential distribution represents the waiting time till the first event, the Gamma distribution represents the waiting time till the rth event. Let X be this waiting time. Then

$$F_X(x) = Pr(X \leq x) = 1 - Pr(\text{fewer than } r \text{ events in interval } [0, x)).$$

The latter is equal to the probability that a Poisson random variable with arrival rate λx is less than r:

$$F_X(x) = 1 - \sum_{k=0}^{r-1} \frac{e^{-\lambda x}(\lambda x)^k}{k!}.$$

Take the derivative with respect to x to get the probability density function:

$$f_X(x) = \sum_{k=0}^{r-1} \frac{\lambda^{k+1} x^k e^{-\lambda x}}{k!} - \sum_{k=1}^{r-1} \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!}.$$

Note that the second summand is only from $k = 1$ to $r - 1$, not from $k = 0$ to $r - 1$. Changing the second summation from $k = 1$ to $k = r - 1$ to the summation from $k = 0$ to $r - 2$, we can write this as

$$f_X(x) = \sum_{k=0}^{r-1} \frac{\lambda^{k+1} x^k e^{-\lambda x}}{k!} - \sum_{k=0}^{r-2} \frac{\lambda^{k+1} x^k e^{-\lambda x}}{k!}$$

$$= \frac{\lambda^r}{(r-1)!} x^{r-1} e^{-\lambda x}.$$

This is a gamma distribution with parameters r and λ.

The gamma distribution is more general. It does not in fact require r to be an integer, although without this it obviously loses the interpretation of waiting for the rth event.

Defining $\beta = 1/\lambda$ and $\alpha = r$, the general pdf for a Gamma distributed random variable X is

$$f_X(x) = \frac{x^{\alpha-1}e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha},$$

for $x > 0$. The parameters α and β are both positive. The gamma function is defined as

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

It has the properties that

$$\Gamma(\alpha+1) = \alpha \cdot \Gamma(\alpha), \qquad \Gamma(1) = \int_0^\infty e^{-t}dt = 1, \qquad \Gamma(1/2) = \sqrt{\pi},$$

implying that

$$\Gamma(n) = (n-1)!,$$

for integer n. Recall that the moment generating function for a random variable Y with an exponential distribution with arrival rate λ is

$$M_Y(t) = E[\exp(tY)] = \frac{\lambda}{\lambda - t} = \frac{1}{1-\beta t},$$

where $\beta = 1/\lambda$ is the mean of the exponential distribution. Let $Y_1, \ldots, Y_\alpha$ be independent exponential random variables with mean β. Then the moment generating function for the gamma distribution is

$$M_X(t) = E\left[\exp\left(\sum_{i=1}^{\alpha} Y_i t\right)\right] = \prod_{i=1}^{\alpha} E[\exp(Y_i t)] = \prod_{i=1}^{\alpha} \frac{1}{1-\beta t} = \frac{1}{(1-\beta t)^\alpha}.$$

The cumulant generating function is

$$K_X(t) = -\alpha \ln(1-\beta t),$$

with

$$K_X'(t) = \frac{\alpha\beta}{1-\beta t}, \qquad K_X''(t) = \frac{\alpha\beta^2}{(1-\beta t)^2}.$$

The mean and variance are $\alpha\beta$ and $\alpha\beta^2$ respectively.

2. Chi–squared Distribution A special case of the Gamma distribution is the Chi–squared distribution. Take $\alpha = k/2$, where k is a positive integer, and $\beta = 2$, we have a Chi–squared distribution with degrees of freedom equal to k. Its pdf is

$$f_X(x) = \frac{x^{k/2-1}e^{-x/2}}{\Gamma(k/2)2^{k/2}},$$

for x positive. Its moment generating function is

$$M_X(t) = \frac{1}{(1-2t)^{k/2}},$$

and its mean and variance are k and $2k$ respectively.

3. Normal Distribution One of the most important distributions is the normal distribution. It does not have as easy a motivation as some of the other distributions, but it is of fundamental importance as an approximation to a large number of statistics through the central limit theorem. A random variable X has a normal distribution with parameters μ and σ^2, denoted by $\mathcal{N}(\mu, \sigma^2)$, if its pdf is

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big),$$

for $-\infty < x < \infty$, and $\sigma^2 > 0$. First consider the moment generating function. Its derivation relies on a trick we have used before, namely using the fact that the pdf and pmf respectively integrate and add up to one:

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(xt - \frac{(x-\mu)^2}{2\sigma^2}\Big)dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-\mu)^2 - 2\sigma^2 xt}{2\sigma^2}\Big)dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(t^2\sigma^2/2 + \mu t - \frac{(x-\mu-\sigma^2 t)^2}{2\sigma^2}\Big)dx \\ &= \exp(\mu t + \sigma^2 t^2/2) \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-\mu-\sigma^2 t)^2}{2\sigma^2}\Big)dx \\ &= \exp(\mu t + \sigma^2 t^2/2). \end{aligned}$$

The cumulant generating function is $K_X(t) = \mu t - \sigma^2 t^2/2$ and hence the mean is μ and the variance σ^2.

One of the most important properties of the normal distribution is that linear transformations of normal random variables are also normally distributed. Consider a random variable X with a $\mathcal{N}(\mu, \sigma^2)$ distribution, and consider the transformation $Y = a + bX$. Then, through the moment generating function,

$$M_Y(t) = E(e^{tY}) = E(e^{(a+bX)t}) = e^{at} \cdot E(e^{(bt)X})$$

$$= e^{ta} \cdot M_X(bt) = \exp(a + \mu bt + \sigma^2 b^2 t^2/2)) = \exp((a + b\mu)t - b^2\sigma^2 t^2/2)$$

$$= \exp(\tilde{\mu} t + \tilde{\sigma}^2 t^2/2).$$

Hence Y has a normal distribution with mean $\tilde{\mu} = a + b\mu$ and variance $\tilde{\sigma}^2 = b^2\sigma^2$. A particularly useful transformation is that from X to $Y = (X - \mu)/\sigma$. Then $Y \sim \mathcal{N}(0, 1)$, the <u>standard normal distribution</u> with mean zero and unit variance.

Finally, there is a close connection between the normal distribution and the chi–squared distribution. If X has a standard normal distribution $\mathcal{N}(0, 1)$, then $Y = X^2$ has a Chi–squared distribution with degrees of freedom equal to one. One argument goes as follows: The moment generating function of Y is

$$M_Y(t) = E(e^{tY}) = E(e^{tX^2}) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{1}{2}x^2 + tx^2\Big) dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2/(1+2t)} x^2\right) dx$$

$$= \sqrt{1/(1-2t)} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi/(1-2t)}} \exp\left(-\frac{1}{2/(1-2t)} x^2\right) dx = \frac{1}{(1-2t)^{1/2}}.$$

This is the mgf for a Chi–squared distribution with degrees of freedom equal to one.

4. <u>Cauchy Distribution</u> A random variable X has a cauchy distribution centered around θ if it has pdf

$$f_X(x) = \frac{1}{\pi} \cdot \frac{1}{1 + (x - \theta)^2},$$

for $-\infty < x < \infty$. This distribution has <u>no</u> moments. The median and mode are both equal to θ. The pdf looks very similar to the pdf for the normal distribution but it has thicker tails. It is often used for modelling variables with high kurtosis, that is, which infrequently take on extremely large values, such as stock prices. An interesting property of the cauchy distribution is that if the sequence of independent random variables $X_1, X_2, \ldots, X_n$ all have the same cauchy distribution centered around θ, then the average $\bar{X} = \sum_{i=1}^{n} X_i/n$ also has that same Cauchy distribution centered around θ, with the same quantiles. In other words, laws of large numbers will be seen not to apply to distributions like Cauchy distributions.

5. <u>Beta Distribution</u>. Suppose two independent random variables Y_1 and Y_2 have Gamma distributions with parameters α and 1 and β and 1. Then the ratio $X = Y_1/(Y_1 + Y_2)$ has a Beta distribution with parameters α and β. The pdf is

$$f_X(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \cdot \Gamma(\beta)} x^{\alpha - 1} \cdot (1 - x)^{\beta - 1},$$

for $0 < x < 1$ and zero elsewhere. The mean and variance are $\alpha/(\alpha + \beta)$ and $\alpha\beta/((\alpha + \beta)^2(\alpha + \beta + 1))$ respectively. This is a very useful distribution for modelling variables that take on values in the unit interval.

6. Let X have a standard normal distribution (that is, a normal distribution with mean zero and unit variance). Let W have a Chi-squared distribution with r degrees of freedom, and let X and W be independent. Then $Y = X/\sqrt{(W/r)}$ has a <u>t-distribution</u> with degrees of freedom equal to r. The probability density function of Y is:

$$f_Y(y) = \frac{\Gamma((r+1)/2)}{\sqrt{\pi r}\Gamma(r/2)} \cdot \frac{1}{(1 + t^2/r)^{(r+1)/2}},$$

for $-\infty < y < \infty$. The t-distribution has thicker tails than the normal distribution. As the degrees of freedom gets larger, then the t-distribution gets closer to the normal distribution. (Note that $E[W/r]=1$, and $V(W/r)=2/r$, so that as $r\to\infty$, the mean of the denominator stays at one, but the variance goes to zero.)

7. Suppose that V and W are independent Chi-squared random variables with degrees of freedom equal to r_1 and r_2 respectively. Then $Y=(V/r_1)/(W/r_2)$ has an F-distribution with degrees of freedom equal to r_1 and r_2. The probability density function is

$$f_Y(y)=\frac{\Gamma((r_1+r_2)/2)(r_1/r_2)^{r_1/2}}{\Gamma(r_1/2)\Gamma(r_2/2)}\cdot\frac{y^{r_1/2-1}}{(1+wr_1/r_2)^{(r_1+r_2)/2}},$$

for positive values of y, and zero elsewhere.

Finally we look at a class of families, the exponential family of distributions.

Definition 1 *A family of pdf's (or pmf's) is an exponential family if it can be written as*

$$f_X(x|\theta)=h(x)\cdot c(\theta)\exp\Big(\sum_{i=1}^{k}w_i(\theta)\cdot t_i(x)\Big),$$

for $-\infty<x<\infty$.

The key feature of this representation is that we can seperate things into functions of the parameters θ and functions of the variable x. If we parametrize this as

$$f_X(x|\eta)=h(x)\cdot c(\eta)\exp\Big(\sum_{i=1}^{k}\eta_i\cdot t_i(x)\Big),$$

for $-\infty<x<\infty$, we refer to η as the natural parameters. Examples include most of the distributions we have looked at sofar:

1. Binomial distribution:

$$f_X(x)=\begin{pmatrix} n \\ x \end{pmatrix}\cdot p^x\cdot(1-p)^{(n-x)},$$

for $x=0,1,2,\ldots,n$, and zero otherwise. This can be written as

$$f_X(x)=1\{x\in\{0,1,2,\ldots,n\}\}\cdot\begin{pmatrix} n \\ x \end{pmatrix}\cdot(1-p)^n\cdot\exp(x\ln p/(1-p)),$$

so, $k = 1$, and

$$h(x) = 1\{x \in \{0, 1, 2, \ldots, n\}\} \cdot \begin{pmatrix} n \\ x \end{pmatrix},$$

$$c(p) = (1 - p)^n,$$

$$w_1(p) = \ln(p/(1 - p)),$$

$$t_1(x) = x.$$

2. Poisson distribution:

$$f_X(x) = 1\{x \in \{1, 2, \ldots\}\} \cdot (1/x!) \cdot \exp(-\lambda)\exp(x \ln \lambda).$$

3. Exponential distribution:

$$f_X(x) = 1\{x > 0\} \cdot \lambda \cdot \exp(-x\lambda).$$

4. Normal distribution:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(\mu^2/(2/\sigma^2)) \cdot \exp(-x^2/(2\sigma^2) + x\mu/\sigma^2).$$

A distribution that does not fit in the exponential familiy is the uniform distribution:

$$f_X(x) = 1\{a < x < b\} \cdot \frac{1}{b - a}.$$

We cannot factor the indicator function $1\{a < x < b\}$ into a function of the parameters and a function of the variable.