STAT/MTHE 353: 3 – Some Special Distributions

T. Linder

Queen's University

Winter 2017

Multinomial Distribution

- Consider an experiment with r possible outcomes such that the probability of the ith outcome is p_i, $i = 1, \ldots, r$, where $p_1 + p_2 + \cdots + p_r = 1$ (generalization of a Bernoulli trial).
- Repeat the experiment *independently* n times and let

$$X_i = \# \text{ of outcomes of type } i \text{ in the } n \text{ trials}$$

- The random variables $(X_1, X_2, \ldots, X_r)$ are said to have a *multinomial distribution* with parameters n and $(p_1, \ldots, p_r)$.
- Note that all the X_i take nonnegative integer values and $X_1 + X_2 + \cdots + X_r = n$.

Joint pmf of multinomial random variables

- Let $x_1, \ldots, x_r \in \mathbb{Z}_+$ such that $x_1 + \cdots + x_r = n$. Then

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_r = x_r) = C_n p_1^{x_1} p_2^{x_2} \cdots p_r^{x_r}$$

  where C_n is the number of sequences of outcomes of length n that have x_1 outcomes of type 1, x_2 outcomes of type 2,..., x_r outcomes of type r.
- Let's use the generalized counting principle: There are $\binom{n}{x_1}$ ways of choosing the x_1 positions for type 1 outcomes. For each such choice, there are $\binom{n-x_1}{x_2}$ ways of choosing the x_2 positions for type 2 outcomes, ... For each choice of the positions of the type $1 \ldots r-1$ objects there are $\binom{n-x_1-\ldots-x_{r-1}}{x_r} = 1$ ways of choosing the x_r positions for type r outcomes.

Thus

$$\begin{aligned} C_n &= \binom{n}{x_1}\binom{n-x_1}{x_2}\cdots\binom{n-x_1-\ldots-x_{r-1}}{x_r} \\ &= \frac{n!}{x_1!x_2!\cdots x_r!} \\ &= \binom{n}{x_1, x_2, \ldots, x_r} \quad \text{(multinomial coefficient)} \end{aligned}$$

We obtain

$$\boxed{P(X_1 = x_1, X_2 = x_2, \ldots, X_r = x_r) = \binom{n}{x_1, x_2, \ldots, x_r} p_1^{x_1} p_2^{x_2} \cdots p_r^{x_r}}$$

for any $x_1, x_2, \ldots, x_r \in \mathbb{Z}_+$ with $x_1 + x_2 + \cdots + x_n = n$.

- Noting that $X_r = n - \sum_{i=1}^{r-1} X_i$, and $p_r = 1 - \sum_{i=1}^{r-1} p_i$ we can equivalently describe the multinomial distribution by the distribution of $(X_1, \ldots, X_{r-1})$:

$$
\begin{aligned}
&P(X_1 = x_1, \ldots, X_{r-1} = x_{r-1}) \\
&\quad = \frac{n!}{x_1! \cdots x_{r-1}! \left(n - \sum_{i=1}^{r-1} x_i\right)!} p_1^{x_1} \cdots p_{r-1}^{x_{r-1}} \left(1 - \sum_{i=1}^{r-1} p_i\right)^{n - \sum_{i=1}^{r-1} x_i}
\end{aligned}
$$

for all $x_1, \ldots, x_{r-1} \in \mathbb{Z}_+$ with $x_1 + \cdots + x_{r-1} \le n$.

Note: For $r = 2$ this is the usual way to write the Binomial(n, p) distribution. In this case $p = p_1$ and $p_2 = 1 - p$.

- The *joint marginal pmfs* can be easily obtained from combinatorial considerations. For $\{i_1, \ldots, i_k\} \subset \{1, \ldots, r\}$ we want the joint pmf of $(X_{i_1}, \ldots, X_{i_k})$. Let's use the common label O for all outcomes *not in* $\{i_1, \ldots, i_k\}$. Thus we have outcomes $i_1, \ldots, i_k$, and O with probabilities $p_{i_1}, \ldots, p_{i_k}$ and $p_O = 1 - p_{i_1} - \cdots - p_{i_k}$.

  Then from the second representation of the multinomial pmf:

$$
\begin{aligned}
&P(X_{i_1} = x_{i_1}, \ldots, X_{i_k} = x_{i_k}) \\
&\quad = \frac{n!}{x_{i_1}! \cdots x_{i_k}! \left(n - \sum_{j=1}^{k} x_{i_j}\right)!} p_{i_1}^{x_{i_1}} \cdots p_{i_k}^{x_{i_k}} \left(1 - \sum_{j=1}^{k} p_{i_j}\right)^{n - \sum_{j=1}^{k} x_{i_j}}
\end{aligned}
$$

  for all $x_{i_1}, \ldots, x_{i_k} \in \mathbb{Z}_+$ with $x_{i_1} + \cdots + x_{i_k} \le n$.

- From this we find that the marginal pdf of X_i is Binomial(n, p_i):

$$P(X_i = x_i) = \frac{n!}{x_i!(n - x_i)!} p_i^{x_i} (1 - p_i)^{n - x_i}, \quad x_i = 0, \ldots, n$$

Gamma Distribution

Definition A continuous r.v. X is said to have a *gamma distribution* with parameters $r > 0$ and $\lambda > 0$ if its pdf is given by

$$\boxed{f(x) = \begin{cases} \dfrac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}, & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases}}$$

where $\Gamma(r)$ is the *gamma function* defined for $r > 0$ by

$$\Gamma(r) = \int_0^\infty y^{r-1} e^{-y} \, dy.$$

Notation: $X \sim \text{Gamma}(r, \lambda)$

Properties of the gamma function

(1) $\Gamma(1/2) = \sqrt{\pi}$.

Proof:

$$
\begin{aligned}
\Gamma(1/2) &= \int_0^\infty \frac{1}{\sqrt{y}} e^{-y} \, dy \qquad \text{(change of variable } y = u^2/2\text{)} \\
&= \int_0^\infty \frac{\sqrt{2}}{u} e^{-u^2/2} u \, du \qquad (dy = u \, du) \\
&= \sqrt{2}\sqrt{2\pi} \underbrace{\int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \, du}_{P(Z>0)=1/2, \text{ where } Z \sim N(0,1)} \\
&= 2\sqrt{\pi}\frac{1}{2} = \sqrt{\pi}. \qquad \square
\end{aligned}
$$

(2) $\Gamma(r) = (r-1)\Gamma(r-1)$ for $r > 1$.

Proof:

$$\begin{aligned}
\Gamma(r) &= \int_0^\infty y^{r-1}e^{-y}\,dy \qquad \text{(integration by parts: } \\
&\qquad\qquad\qquad u = y^{r-1},\ dv = e^{-y}\,dy) \\
&= \left[-y^{r-1}e^{-y}\right]_0^\infty + \int_0^\infty (r-1)y^{r-2}e^{-y}\,dy \\
&= (r-1)\int_0^\infty y^{r-2}e^{-y}\,dy \\
&= (r-1)\Gamma(r-1). \qquad \square
\end{aligned}$$

Corollary: If r is a positive integer, then $\Gamma(r) = (r-1)!$

Proof: Noting that $\Gamma(1) = \int_0^\infty e^{-y}\,dy = 1$,

$$\begin{aligned}
\Gamma(r) &= (r-1)\Gamma(r-1) = (r-1)(r-2)\Gamma(r-2) = \ldots \\
&= (r-1)(r-2)\cdots 2\cdot 1\cdot \Gamma(1) = (r-1)! \qquad \square
\end{aligned}$$

Moments $E(X^k)$: For $X \sim \text{Gamma}(r,\lambda)$ and $k \geq 1$ an integer,

$$\begin{aligned}
E(X^k) &= \int_0^\infty x^k\left(\frac{\lambda^r}{\Gamma(r)}x^{r-1}e^{-\lambda x}\right)dx \\
&= \frac{\lambda^r}{\Gamma(r)}\int_0^\infty x^{r+k-1}e^{-\lambda x}\,dx \\
&= \frac{\lambda^r}{\Gamma(r)}\frac{\Gamma(r+k)}{\lambda^{r+k}}\int_0^\infty \frac{\lambda^{r+k}}{\Gamma(r+k)}x^{r+k-1}e^{-\lambda x}\,dx \\
&= \frac{\Gamma(r+k)}{\Gamma(r)\lambda^k} = \frac{(r+k-1)\Gamma(r+k-1)}{\Gamma(r)\lambda^k} \\
&= \frac{(r+k-1)(r+k-2)\cdots r\Gamma(r)}{\Gamma(r)\lambda^k} \\
&= \boxed{\frac{(r+k-1)(r+k-2)\cdots r}{\lambda^k}}
\end{aligned}$$

For $k=1$ we get $\boxed{E(X) = \frac{r}{\lambda}}$; for $k=2$, $\boxed{E(X^2) = \frac{(r+1)r}{\lambda^2}}$, so

$$\text{Var}(X) = \frac{(r+1)r}{\lambda^2} - \left(\frac{r}{\lambda}\right)^2 = \boxed{\frac{r}{\lambda^2}}$$

Special Cases

- If $r=1$, then $f(x) = \lambda e^{-\lambda x}$, $x > 0$, so $X \sim \text{Exp}(\lambda)$, i.e, X has the *exponential distribution* with parameter λ. Thus

$$\text{Exp}(\lambda) = \text{Gamma}(1,\lambda)$$

- If $r = k/2$ for some positive integer k and $\lambda = 1/2$, then

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

  This is called the χ^2 (chi-squared) distribution with k degrees of freedom (χ^2_k).

Example: ...

Beta function

Let $\alpha, \beta > 0$ and consider

$$\begin{aligned}
\Gamma(\alpha)\Gamma(\beta) &= \left(\int_0^\infty x^{\alpha-1}e^{-x}\,dx\right)\left(\int_0^\infty y^{\beta-1}e^{-y}\,dy\right) \\
&= \int_0^\infty\int_0^\infty x^{\alpha-1}y^{\beta-1}e^{-(x+y)}\,dxdy.
\end{aligned}$$

Use change of variables $u = x+y$, $v = x/(x+y)$ with inverse

$$x = uv, \qquad y = u - uv = (1-v)u.$$

The region $\{x > 0, y > 0\}$ is mapped onto $\{u > 0, 0 < v < 1\}$. The Jacobian of the inverse is

$$J(u,v) = \det\begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v}\end{bmatrix} = \det\begin{bmatrix} v & u \\ 1-v & -u\end{bmatrix} = -vu - (1-v)u = -u$$

We obtain

$$\begin{aligned}
\Gamma(\alpha)\Gamma(\beta) &= \int_0^\infty \int_0^\infty x^{\alpha-1} y^{\beta-1} e^{-(x+y)}\, dxdy \\
&= \int_0^1 \int_0^\infty (uv)^{\alpha-1} (u(1-v))^{\beta-1} e^{-u} |-u|\, dudv \\
&= \int_0^1 \int_0^\infty u^{\alpha+\beta-1} e^{-u} v^{\alpha-1} (1-v))^{\beta-1}\, dudv \\
&= \Bigg(\underbrace{\int_0^\infty u^{\alpha+\beta-1} e^{-u}\, du}_{\Gamma(\alpha+\beta)}\Bigg)\Bigg(\int_0^1 v^{\alpha-1}(1-v)^{\beta-1}\, dv\Bigg)
\end{aligned}$$

Define the *beta function* of two positive arguments α and β by

$$\boxed{B(\alpha,\beta) = \int_0^1 v^{\alpha-1}(1-v)^{\beta-1}\, dv}$$

We have obtained

$$\boxed{B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}}$$

Example: Suppose $X_1 \sim \text{Gamma}(r_1, \lambda)$ and $X_2 \sim \text{Gamma}(r_2, \lambda)$ are independent. Find the pdf of $U = X_1 + X_2$.

Solution: ...

Conclusion: the family of gamma distributions with given λ *is closed* under sums of independent random variables.

- We have seen that if $X_1 \sim \text{Gamma}(r_1, \lambda)$ and $X_2 \sim \text{Gamma}(r_2, \lambda)$ are independent, then $X_1 + X_2 \sim \text{Gamma}(r_1 + r_2, \lambda)$.
- *Inductively*, if $X_1, \ldots, X_n$ are independent with $X_i \sim \text{Gamma}(r_i, \lambda)$, then

$$X_1 + \cdots + X_n \sim \text{Gamma}(r_1 + \cdots + r_n, \lambda)$$

- Also, we saw that if $Z \sim N(0,1)$, then $Z^2 \sim \text{Gamma}(1/2, 1/2)$ (i.e., $Z^2 \sim \chi^2_1$).
- Combining the above gives that if $Z_1, \ldots, Z_n$ are i.i.d. $N(0,1)$ random variables, then

$$\boxed{Z_1^2 + \cdots + Z_n^2 \sim \text{Gamma}(n/2, 1/2) = \chi^2_n}$$

- This result is often used in statistics.

Let $Z_1, \ldots, Z_n$ be i.i.d. random variables with common mean μ and variance σ^2. The *sample mean* and *sample variance* are defined by

$$\bar{Z} = \frac{1}{n}\sum_{i=1}^n Z_i, \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^n (Z_i - \bar{Z})^2$$

Example: Show that $E(\bar{Z}) = \mu$ and $E(S^2) = \sigma^2$.

An important result in statistics is the following:

Lemma 1

Assume $Z_1, \ldots, Z_n$ are i.i.d. $N(0,1)$. Then

$$\bar{Z} \sim N(0, 1/n), \qquad (n-1)S^2 \sim \chi^2_{n-1}$$

and $\bar{Z}$ and S^2 are independent.

Before proving the lemma, let's review a few facts about orthogonal (linear) transformations on $\mathbb{R}^n$.

- An $n \times n$ real matrix $\boldsymbol{A}$ is called *orthogonal* if $\boldsymbol{A}^T = \boldsymbol{A}^{-1}$, i.e., $\boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{I}$ (the $n \times n$ identity matrix).
- An orthogonal $\boldsymbol{A}$ does not change the norm (length) of its argument:

$$\sum_{i=1}^n x_i^2 = \|\boldsymbol{x}\|^2 = \boldsymbol{x}^T\boldsymbol{x} = \boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = (\boldsymbol{A}\boldsymbol{x})^T(\boldsymbol{A}\boldsymbol{x}) = \|\boldsymbol{A}\boldsymbol{x}\|^2$$

- If $\boldsymbol{A}$ is orthogonal, then $|\det \boldsymbol{A}| = 1$.

Now let $\boldsymbol{Z} = (Z_1, \ldots, Z_n)$ have joint pdf $f(\boldsymbol{z})$. Letting $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{Z}$ for an orthogonal $\boldsymbol{A}$, we have $\boldsymbol{Z} = \boldsymbol{A}^{-1}\boldsymbol{Y}$. By the transformation formula the pdf $f_{\boldsymbol{Y}}(\boldsymbol{y})$ of $\boldsymbol{Y}$ is

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = f(\boldsymbol{A}^{-1}\boldsymbol{y})|J| = f(\boldsymbol{A}^{-1}\boldsymbol{y})\frac{1}{|\det(\boldsymbol{A})|} = f(\boldsymbol{A}^T\boldsymbol{y}).$$

Proof of Lemma: The joint pdf of $Z_1, \ldots, Z_n$ is

$$f(\boldsymbol{z}) = f(z_1, \ldots, z_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-z_i^2/2} = \Big(\frac{1}{\sqrt{2\pi}}\Big)^n e^{-\frac{1}{2}\sum_{i=1}^n z_i^2}$$

Let $\boldsymbol{A}$ be an $n \times n$ matrix with first row equal to $(1/\sqrt{n}, \ldots, 1/\sqrt{n})$ and choose rows $2, \ldots, n$ in any way so that they have unit length and they are orthogonal to all other rows. $\boldsymbol{A}$ constructed this way is *orthogonal.*

The joint pdf of $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{Z}$ is

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = f(\boldsymbol{A}^T\boldsymbol{y}) = \Big(\frac{1}{\sqrt{2\pi}}\Big)^n e^{-\frac{1}{2}\sum_{i=1}^n y_i^2}$$

since $\boldsymbol{A}^T$ is orthogonal and so $\|\boldsymbol{A}^T\boldsymbol{y}\|^2 = \|\boldsymbol{y}\|^2 = \sum_{i=1}^n y_i^2$.

Thus $Y_1, \ldots, Y_n$ are i.i.d. $N(0,1)$.

Proof cont'd: From $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{Z}$, we have

$$Y_1 = \frac{1}{\sqrt{n}}\sum_{i=1}^n Z_i = \frac{n\bar{Z}}{\sqrt{n}} = \sqrt{n}\bar{Z}$$

and

$$\begin{aligned} Y_2^2 + \cdots + Y_n^2 &= \Big(\sum_{i=1}^n Y_i^2\Big) - Y_1^2 = \Big(\sum_{i=1}^n Z_i^2\Big) - n\bar{Z}^2 \\ &= \ldots = \sum_{i=1}^n (Z_i - \bar{Z})^2 = (n-1)S^2. \end{aligned}$$

Since $\bar{Z}$ is a function of Y_1, S^2 is a function of $Y_2, \ldots, Y_n$, we get that $\bar{Z}$ and S^2 are independent (since $Y_1, Y_2, \ldots, Y_n$ are independent).

Since $\bar{Z} = Y_i/\sqrt{n}$ and $Y_i \sim N(0,1)$, we obtain $\bar{Z} \sim N(0, 1/n)$.

Since $(n-1)S^2 = Y_2^2 + \cdots + Y_n^2$, we have $(n-1)S^2 \sim \text{Gamma}((n-1)/2, 1/2) = \chi^2_{n-1}$. □

Connection with Poisson Process

- Recall: If X_1 denotes the time of the *first* event occurring in a Poisson process with rate λ, then $X_1 \sim \text{Exp}(\lambda)$.
- The following can be shown: For $i = 1, 2, \ldots, n$ let X_i denote the time between the occurrence of the $(i-1)$th and the ith events in a Poisson process with rate λ. Then the random variables $X_1, \ldots, X_n$ are independent and $X_i \sim \text{Exp}(\lambda)$.
- Let $S_n = X_1 + \cdots + X_n$, the time till until the nth event. Since $\text{Exp}(\lambda) = \text{Gamma}(1, \lambda)$, we obtain that

$$\boxed{S_n \sim \text{Gamma}(n, \lambda).}$$

Beta Distribution

Definition A continuous r.v. X is said to have a *beta distribution* with parameters $\alpha > 0$ and $\beta > 0$ if its pdf is given by

$$\boxed{f(x) = \begin{cases} \dfrac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, & \text{if } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}}$$

where the beta function $B(\alpha, \beta)$ is given by

$$B(\alpha,\beta) = \int_0^1 y^{\alpha-1}(1-y)^{\beta-1}\,dy = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

Notation: $X \sim \text{Beta}(\alpha, \beta)$

Moments $E(X^k)$: For $X \sim \text{Beta}(\alpha,\beta)$ and $k \geq 1$ an integer,

$$\begin{aligned} E(X^k) &= \int_0^1 x^k \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}\,dx \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \underbrace{\int_0^1 x^{k+\alpha-1}(1-x)^{\beta-1}\,dx}_{B(k+\alpha,\beta)} \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(k+\alpha)\Gamma(\beta)}{\Gamma(k+\alpha+\beta)} \\ &= \frac{(k+\alpha-1)\cdots\alpha}{(k+\alpha+\beta-1)\cdots(\alpha+\beta)} \end{aligned}$$

Letting $k = 1, 2$ we get $\boxed{E(X) = \frac{\alpha}{\alpha+\beta}}$ and $\boxed{E(X^2) = \frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)}}$, so

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

Examples:

- For $\alpha = \beta = 1$ we obtain $X \sim \text{Uniform}(0,1)$ having mean $1/2$ and variance $1/12$.
- Recall that the pdf of the kth order statistics $X_{(k)}$ of random sample $X_1, \ldots, X_n$ with common cdf $F(x)$ is

  $$f_k(x) = \frac{n!}{(k-1)!(n-k)!} f(x) F(x)^{k-1} \big(1 - F(x)\big)^{n-k}$$

  If the X_i are sampled from Uniform$(0,1)$, then $F(x) = x$ for $0 < x < 1$ and we get

  $$f_k(x) = \begin{cases} \frac{n!}{(k-1)!(n-k)!} x^{k-1}(1-x)^{n-k} & \text{if } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

  Thus $X_{(k)} \sim \text{Beta}(k, n-k+1)$.

The beta distribution is useful as a model for random variables that take values in a bounded interval, say (a, b).

Example: Let $X \sim \text{Beta}(\alpha,\beta)$ and let $Y = (b-a)X + a$. Find the pdf of Y.

Solution: ...

Example: (Connection with gamma distribution) Assume $X_1, \ldots, X_n$ are independent with $X_i \sim \text{Gamma}(r_i, \alpha)$. Show that

$$\frac{X_i}{\sum_{j=1}^n X_j} \sim \text{Beta}(r_i, r_{-i})$$

where $r_{-i} = \left(\sum_{j=1}^n r_j\right) - r_i$.

Solution: ...