Other Common Univariate Distributions

Dear students, besides the Normal, Bernoulli and Binomial distributions, the following distributions are also very important in our studies.

1. Discrete Distributions

1.1. Geometric distribution

This is a discrete waiting time distribution. Suppose a sequence of independent Bernoulli trials is performed and let X be the number of failures preceding the first success. Then \( X \sim \text{Geom}(p) \), with pdf

\[
f(x) = p(1-p)^x, x = 0,1,2, ...
\]

1.2. Negative Binomial distribution

Suppose a sequence of independent Bernoulli trials is conducted. If X is the number of failures preceding the nth success, then X has a negative binomial distribution.

Probability density function:

\[
f(x) = \binom{n+x-1}{n-1} p^n (1-p)^x,
\]

\[
E(X) = \frac{n(1-p)}{p},
\]

\[
Var(X) = \frac{n(1-p)}{p^2},
\]

\[
M_X(t) = \left[ \frac{p}{1 - e^t (1-p)} \right]^n.
\]

1. If \( n = 1 \), we obtain the geometric distribution.

2. Also seen to arise as sum of \( n \) independent geometric variables.

1.3. Poisson distribution

Parameter: rate \( \lambda > 0 \)

MGF: \( M_X(t) = e^{\lambda(e^t-1)} \)

Probability density function:
1. The Poisson distribution arises as the distribution for the number of “point events” observed from a Poisson process.

Examples:

![Poisson Example](image)

2. The Poisson distribution also arises as the limiting form of the binomial distribution:

\[ n \to \infty, \quad np \to \lambda \]

\[ p \to 0 \]

The derivation of the Poisson distribution (via the binomial) is underpinned by a Poisson process i.e., a point process on \([0, \infty)\); see Figure 1.

**AXIOMS** for a Poisson process of rate \( \lambda > 0 \) are (That is, a counting process is a Poisson process if it satisfies the following rules):

**(A)** The number of occurrences in disjoint intervals are independent.

**(B)** Probability of exactly 1 occurrence in any sub-interval \([t, t + h)\) is \(\lambda h + o(h)\) (approx prob. is equal to length of interval (h) times \(\lambda\)).

**(C)** Probability of more than one occurrence in \([t, t + h)\) is \(o(h)\) (h \(\to 0\)) (i.e. prob is small, negligible).

**Note:** \(o(h)\), pronounced (small order \(h\)) is standard notation for any function \(r(h)\) with the property:

\[ \lim_{h \to 0} \frac{r(h)}{h} = 0 \]

It can be derived that for a Poisson process with rate \(\lambda\), the number of events occurring in an interval with length \(t\), denoted as \(N(t)\) would follow a Poisson distribution with parameter \(\lambda t\). That is,
\[ P(N(t) = k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!} \], for \( k = 0, 1, 2, \ldots \)

One can also show that the **inter-arrival time** follows iid exponential(\( \lambda \)) distribution, and subsequently, the **waiting time**, being the summation of iid exponential variables, follows the gamma distribution.

**Poisson distribution examples**

1. The number of road construction projects that take place at any one time in a certain city follows a Poisson distribution with a mean of 3. Find the probability that exactly five road construction projects are currently taking place in this city. (0.100819)

2. The number of road construction projects that take place at any one time in a certain city follows a Poisson distribution with a mean of 7. Find the probability that more than four road construction projects are currently taking place in the city. (0.827008)

3. The number of traffic accidents that occur on a particular stretch of road during a month follows a Poisson distribution with a mean of 7.6. Find the probability that less than three accidents will occur next month on this stretch of road. (0.018757)

4. The number of traffic accidents that occur on a particular stretch of road during a month follows a Poisson distribution with a mean of 7. Find the probability of observing exactly three accidents on this stretch of road next month. (0.052129)

5. The number of traffic accidents that occur on a particular stretch of road during a month follows a Poisson distribution with a mean of 6.8. Find the probability that the next two months will both result in four accidents each occurring on this stretch of road. (0.009846)

6. Suppose the number of babies born during an 8-hour shift at a hospital’s maternity wing follows a Poisson distribution with a mean of 6 an hour. Find the probability that five babies are born during a particular 1-hour period in this maternity wing. (0.160623)

7. The university policy department must write, on average, five tickets per day to keep department revenues at budgeted levels. Suppose the number of tickets written per day follows a Poisson distribution with a mean of 8.8 tickets per day. Find the probability that less than six tickets are written on a randomly
selected day from this distribution. (0.128387)

8. The number of goals scored at State College hockey games follows a Poisson distribution with a mean of 3 goals per game. Find the probability that each of four randomly selected State College hockey games resulted in six goals being scored. (.00000546)

1.4. Hypergeometric distribution

Consider an urn containing $M$ black and $N$ white balls. Suppose $n$ balls are sampled randomly without replacement and let $X$ be the number of black balls chosen. Then $X$ has a hypergeometric distribution.

**Parameters:** $M, N > 0$, $0 < n \leq M + N$

**Possible values:** $\text{max}(0, n - N) \leq x \leq \text{min}(n, M)$

**Prob. density function:**

$$f(x) = \binom{M}{x} \binom{N}{n-x} \binom{M+N}{n},$$

$$E(X) = n \frac{M}{M+N}, \quad \text{Var}(x) = \frac{M+N-n}{M+N-1} \frac{n MN}{(M+N)^2}.$$ 

The mgf exists, but there is no useful expression available.

1. The hypergeometric PDF is simply

$$\frac{\text{# samples with } x \text{ black balls}}{\text{# possible samples}} = \binom{M}{x} \binom{N}{n-x} \binom{M+N}{n},$$

2. To see how the limits arise, observe we must have $x \leq n$ (i.e., no more than sample size of black balls in the sample.) Also, $x \leq M$, i.e., $x \leq \text{min}(n, M)$.

Similarly, we must have $x \geq 0$ (i.e., cannot have $< 0$ black balls in sample), and

$n - x \leq N$ (i.e., cannot have more white balls than number in urn).

i.e. $x \geq n - N$

i.e. $x \geq \text{max}(0, n - N)$.

3. If we sample with replacement, we would get $X \sim B\left(n, p = \frac{M}{M+N}\right)$. It is interesting to compare moments:
<table>
<thead>
<tr>
<th>Distribution</th>
<th>$E(X) = np$</th>
<th>$Var(X) = \frac{M + N - n}{M + N - 1} [np(1 - p)]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypergeometric</td>
<td>$E(x) = np$</td>
<td>$Var(X) = np(1 - p)$</td>
</tr>
</tbody>
</table>
2. Continuous Distributions

2.1 Uniform Distribution

For $X \sim U(a, b)$, its pdf and cdf are:

$$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{otherwise} \end{cases}$$

$$F(x) = \int_{-\infty}^{x} f(x) \, dx = \int_{a}^{x} \frac{1}{b-a} \, dx = \frac{x-a}{b-a}, \quad \text{for } a < x < b$$

The more complete form of the cdf is:

$$F(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & b \leq x \end{cases}$$

Its mgf is:

$$M_X(t) = \frac{e^{tb} - e^{ta}}{t(b-a)}$$

A special case is the $X \sim U(0, 1)$ distribution:

$$f(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$F(x) = x \text{ for } 0 < x < 1$$

$$E(X) = \frac{1}{2}, \quad Var(X) = \frac{1}{12}, \quad M(t) = \frac{e^t - 1}{t}.$$

2.2 Exponential Distribution

**PDF:** $f(x) = \lambda e^{-\lambda x}, x \geq 0$

**CDF:** $F(x) = 1 - e^{-\lambda x}$,

**MGF:**

$$M_X(t) = \frac{\lambda}{\lambda - t}, \quad \lambda > 0, x \geq 0.$$ 

This is the distribution for the waiting time until the first occurrence in a Poisson process with rate parameter $\lambda > 0$.

1. If $X \sim Exp(\lambda)$ then,

$$P(X \geq t + x | X \geq t) = P(X \geq x)$$

(memoryless property)
2. It can be obtained as limiting form of geometric distribution.
3. One can also easily derive that the geometric distribution also has the memoryless property.

### 2.3 Gamma distribution

Given $X \sim \text{gamma}(\alpha, \beta)$

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1}e^{-\frac{x}{\beta}}, \ 0 < x < \infty$$

Or, some books use $\lambda = \frac{1}{\beta}$

Then:

$$f(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1}e^{-\lambda x}, \ 0 < x < \infty$$

If $r$ is a non-negative integer, then $\Gamma(r) = (r-1)!$

$$1 = \int_{-\infty}^{\infty} f(x)\,dx = \int_{0}^{\infty} \frac{\lambda^r}{\Gamma(r)} x^{r-1}e^{-\lambda x}\,dx$$

$$\Gamma(r) = \lambda^r \int_{0}^{\infty} x^{r-1}e^{-\lambda x}\,dx$$

$$M_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-r} = \left(\frac{\lambda-t}{\lambda}\right)^{-r}$$

$$\Rightarrow M_X(t) = \left(\frac{\lambda}{\lambda-t}\right)^r$$

$$E(X) = \frac{r}{\lambda}$$

$$\text{Var}(X) = \frac{r}{\lambda^2}$$

**Special case** : when $\frac{r}{2}, \ \frac{\lambda}{2} \Rightarrow X \sim \chi^2_k$

**Special case** : when $r = 1 \Rightarrow X \sim \exp(\lambda)$
Review

\( X \sim \exp(\lambda) \)

p.d.f. \( f(x) = \lambda e^{-\lambda x}, \; x \succ 0 \)

m.g.f. \( M_X(t) = \frac{\lambda}{\lambda - t} \)

e.g. Let \( X_i \sim \exp(\lambda), \; i = 1, \ldots, n \). What is the distribution of \( \sum_{i=1}^{n} X_i \) ?

Solution

\[ M_{\sum_{i=1}^{n} X_i}(t) = \prod_{i=1}^{n} M_{X_i}(t) = \left( \frac{\lambda}{\lambda - t} \right)^n \]

\( \therefore \sum_{i=1}^{n} X_i \sim \text{gamma}(r = n, \lambda) \)

e.g. Let \( W \sim \chi_k^2 \). What is the mgf of \( W \) ?

Solution

\[ M_W(t) = \left( \frac{1}{2} \right)^{\frac{k}{2}} \left( \frac{1}{2 - t} \right)^{\frac{k}{2}} \]

\[ \Rightarrow M_W(t) = \left( \frac{1}{1 - 2t} \right)^{\frac{k}{2}} \]

e.g. Let \( W_1 \sim \chi_{k_1}^2, \; W_2 \sim \chi_{k_2}^2, \) and \( W_1 \) and \( W_2 \) are independent. What is the distribution of \( W_1 + W_2 \) ?

Solution

\[ M_{W_1 + W_2}(t) = M_{W_1}(t) \cdot M_{W_2}(t) = \left( \frac{1}{1 - 2t} \right)^{\frac{k_1}{2}} \left( \frac{1}{1 - 2t} \right)^{\frac{k_2}{2}} = \left( \frac{1}{1 - 2t} \right)^{\frac{k_1 + k_2}{2}} \]

\( \Rightarrow W_1 + W_2 \sim \chi_{k_1 + k_2}^2 \)

*** In general, for \( X \sim \text{gamma}(\alpha, \lambda), \; \alpha > 0 \).

(Note, sometimes we use \( r \) instead of \( \alpha \) as shown above.)
1. $\alpha$ is the shape parameter, $\lambda$ is the scale parameter.

Note: if $Y \sim \text{gamma}(\alpha, 1)$ and $X = \frac{Y}{\lambda}$, then $X \sim \text{gamma}(\alpha, \lambda)$. That is, $\lambda$ is the scale parameter.

Figure 2: Gamma Distribution

2. Gamma $\left( \frac{k}{2}, \frac{1}{2} \right)$ distribution is also called $\chi_k^2$ (chi-square with $k$ df) distribution, if $k$ is a positive integer;

3. The $\text{gamma}(K, \lambda)$ random variable can also be interpreted as the waiting time until the $K^{th}$ occurrence (events) in a Poisson process.

### 2.4 Beta density function

Suppose $Y_1 \sim \text{Gamma}(\alpha, \lambda)$, $Y_2 \sim \text{Gamma}(\beta, \lambda)$ independently, then,

$$X = \frac{Y_1}{Y_1 + Y_2} \sim B(\alpha, \beta), 0 \leq x \leq 1.$$ 

Remark: we have derived this in Lecture 5 – transformation. I have copied it here again, as a review.

e.g. Suppose that $Y_1 \sim \text{gamma}(\alpha, 1)$, $Y_2 \sim \text{gamma}(\beta, 1)$, and that $Y_1$ and $Y_2$ are independent. Define the transformation

$$U_1 = g_1(Y_1, Y_2) = Y_1 + Y_2$$

$$U_2 = g_2(Y_1, Y_2) = \frac{Y_1}{Y_1 + Y_2}.$$ 

Find each of the following distributions:

(a) $f_{U_1,U_2}(u_1, u_2)$, the joint distribution of $U_1$ and $U_2$,
(b) $f_{U_1}(u_1)$, the marginal distribution of $U_1$, and
(c) $f_{U_2}(u_2)$, the marginal distribution of $U_2$.

Solutions. (a) Since $Y_1$ and $Y_2$ are independent, the joint distribution of $Y_1$ and $Y_2$ is

$$f_{Y_1,Y_2}(y_1, y_2) = f_{Y_1}(y_1)f_{Y_2}(y_2)$$
\[
\begin{align*}
&= \frac{1}{\Gamma(\alpha)} y_1^{\alpha-1} e^{-y_1} \times \frac{1}{\Gamma(\beta)} y_2^{\beta-1} e^{-y_2} \\
&= \frac{1}{\Gamma(\alpha)\Gamma(\beta)} y_1^{\alpha-1} y_2^{\beta-1} e^{-(y_1+y_2)},
\end{align*}
\]

for \( y_1 > 0, y_2 > 0 \), and 0, otherwise. Here, \( R_{Y_1,Y_2} = \{(y_1, y_2) : y_1 > 0, y_2 > 0\} \). By inspection, we see that \( u_1 = y_1 + y_2 > 0 \), and \( u_2 = \frac{y_1}{y_1+y_2} \) must fall between 0 and 1.

Thus, the domain of \( U = (U_1, U_2) \) is given by
\[ R_{U_1,U_2} = \{(u_1, u_2) : u_1 > 0, 0 < u_2 < 1\}. \]

The next step is to derive the inverse transformation. It follows that
\[
\begin{align*}
u_1 &= y_1 + y_2 \\
u_2 &= \frac{y_1}{y_1 + y_2}
\end{align*}
\]

and
\[
\begin{align*}
y_1 &= g_1^{-1}(u_1, u_2) = u_1 u_2 \\
y_2 &= g_2^{-1}(u_1, u_2) = u_1 - u_1 u_2
\end{align*}
\]

The Jacobian is given by
\[
J = \det \begin{vmatrix} \frac{\partial g_1^{-1}(u_1, u_2)}{\partial u_1} & \frac{\partial g_1^{-1}(u_1, u_2)}{\partial u_2} \\ \frac{\partial g_2^{-1}(u_1, u_2)}{\partial u_1} & \frac{\partial g_2^{-1}(u_1, u_2)}{\partial u_2} \end{vmatrix} = \det \begin{vmatrix} u_2 & u_1 \\ 1 - u_2 & -u_1 \end{vmatrix} = -u_1 u_2 - u_1 (1 - u_2) = -u_1.
\]

We now write the joint distribution for \( U = (U_1, U_2) \). For \( u_1 > 0 \) and \( 0 < u_2 < 1 \), we have that
\[
f_{U_1, U_2}(u_1, u_2) = f_{Y_1, Y_2}[g_1^{-1}(u_1, u_2), g_2^{-1}(u_1, u_2)] |J| = \frac{1}{\Gamma(\alpha)\Gamma(\beta)} (u_1 u_2)^{\alpha-1} (u_1 - u_1 u_2)^{\beta-1} e^{-(u_1 u_2 + (u_1 - u_1 u_2))} \times | - u_1|
\]

Note: We see that \( U_1 \) and \( U_2 \) are independent since the domain \( R_{U_1,U_2} = \{(u_1, u_2) : u_1 > 0, 0 < u_2 < 1\} \) does not constrain \( u_1 \) by \( u_2 \) or vice versa and since the nonzero part of \( f_{U_1, U_2}(u_1, u_2) \) can be factored into the two expressions \( h_1(u_1) \) and \( h_2(u_2) \), where
\[
h_1(u_1) = u_1^{\alpha+\beta-1} e^{-u_1}
\]
and
\[
h_2(u_2) = \frac{u_2^{\alpha-1}(1 - u_2)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)}.
\]

(b) To obtain the marginal distribution of \( U_1 \), we integrate the joint pdf \( f_{U_1, U_2}(u_1, u_2) \) over \( u_2 \). That is, for \( u_1 > 0 \),
\[
f_{U_1}(u_1) = \int_{u_2=0}^{1} f_{U_1, U_2}(u_1, u_2) \, du_2
\]
\[
\int_{u_2=0}^{1} u_2^{\alpha-1} (1-u_2)^{\beta-1} \frac{u_1^{\alpha+\beta-1} e^{-u_1}}{\Gamma(\alpha)\Gamma(\beta)} du_2
\]
\[
= \frac{1}{\Gamma(\alpha)\Gamma(\beta)} \int_{u_2=0}^{1} u_2^{\alpha-1} (1-u_2)^{\beta-1} e^{-u_1} du_2
\]
\[
= \frac{1}{\Gamma(\alpha+\beta)} u_1^{\alpha+\beta-1} e^{-u_1}
\]

Summarizing,
\[
f_{U_1}(u_1) = \begin{cases} 
\frac{1}{\Gamma(\alpha+\beta)} u_1^{\alpha+\beta-1} e^{-u_1}, & u_1 > 0 \\
0, & \text{otherwise.} 
\end{cases}
\]

We recognize this as a \(\text{gamma}(\alpha + \beta, 1)\) pdf; thus, marginally, \(U_1 \sim \text{gamma}(\alpha + \beta, 1)\).

(c) To obtain the marginal distribution of \(U_2\), we integrate the joint pdf \(f_{U_1,U_2}(u_1,u_2)\) over \(u_2\). That is, for \(0 < u_2 < 1\),
\[
f_{U_2}(u_2) = \int_{u_1=0}^{\infty} f_{U_1,U_2}(u_1,u_2) du_1
\]
\[
= \int_{u_1=0}^{\infty} \frac{u_2^{\alpha-1} (1-u_2)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)} u_1^{\alpha+\beta-1} e^{-u_1} du_1
\]
\[
= \frac{u_2^{\alpha-1} (1-u_2)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)} \int_{u_1=0}^{\infty} u_1^{\alpha+\beta-1} e^{-u_1} du_1
\]
\[
= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} u_2^{\alpha-1} (1-u_2)^{\beta-1}.
\]

Summarizing,
\[
f_{U_2}(u_2) = \begin{cases} 
\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} u_2^{\alpha-1} (1-u_2)^{\beta-1}, & 0 < u_2 < 1 \\
0, & \text{otherwise.} 
\end{cases}
\]

Thus, marginally, \(U_2 \sim \text{beta}(\alpha, \beta)\). □
2.5 Standard Cauchy distribution

Possible values: \( x \in R \)

PDF: \( f(x) = \frac{1}{\pi} \left( \frac{1}{1+x^2} \right) \); (location parameter \( \theta = 0 \))

CDF: \( F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan x \)

\( E(X), Var(X), M_X(t) \) do not exist.

The Cauchy is a bell-shaped distribution symmetric about zero for which no moments are defined.

If \( Z_1 \sim N(0, 1) \) and \( Z_2 \sim N(0, 1) \) independently, then \( X = \frac{Z_1}{Z_2} \sim \text{Cauchy distribution} \). Please also see the lecture notes on transformation for proof (Lecture 5).