Stat 5101 Notes: Brand Name Distributions

Charles J. Geyer

February 25, 2019

Contents
1 Discrete Uniform Distribution 2
2 General Discrete Uniform Distribution 2
3 Uniform Distribution 3
4 General Uniform Distribution 3
5 Bernoulli Distribution 4
6 Binomial Distribution 5
7 Hypergeometric Distribution 6
8 Poisson Distribution 7
9 Geometric Distribution 8
10 Negative Binomial Distribution 9
11 Normal Distribution 10
12 Exponential Distribution 12
13 Gamma Distribution 12
14 Beta Distribution 14
15 Multinomial Distribution 15
1 Discrete Uniform Distribution

Abbreviation  DiscUnif(n).

Type  Discrete.

Rationale  Equally likely outcomes.

Sample Space  The interval 1, 2, . . . , n of the integers.

Probability Mass Function

\[ f(x) = \frac{1}{n}, \quad x = 1, 2, \ldots, n \]

Moments

\[ E(X) = \frac{n + 1}{2} \]
\[ \text{var}(X) = \frac{n^2 - 1}{12} \]

2 General Discrete Uniform Distribution

Type  Discrete.
**Sample Space**  Any finite set $S$.

**Probability Mass Function**

$$f(x) = \frac{1}{n}, \quad x \in S,$$

where $n$ is the number of elements of $S$.

### 3 Uniform Distribution

**Abbreviation**  Unif($a, b$).

**Type**  Continuous.

**Rationale**  Continuous analog of the discrete uniform distribution.

**Parameters**  Real numbers $a$ and $b$ with $a < b$.

**Sample Space**  The interval $(a, b)$ of the real numbers.

**Probability Density Function**

$$f(x) = \frac{1}{b-a}, \quad a < x < b$$

**Moments**

$$E(X) = \frac{a+b}{2}$$

$$\text{var}(X) = \frac{(b-a)^2}{12}$$

**Relation to Other Distributions**  Beta(1, 1) = Unif(0, 1).

### 4 General Uniform Distribution

**Type**  Continuous.

**Sample Space**  Any open set $S$ in $\mathbb{R}^n$. 

3
5 Bernoulli Distribution

Abbreviation Ber(p).

Type Discrete.

Rationale Any zero-or-one-valued random variable.

Parameter Real number $0 \leq p \leq 1$.

Sample Space The two-element set \{0,1\}.

Probability Mass Function

$$f(x) = \begin{cases} p, & x = 1 \\ 1 - p, & x = 0 \end{cases}$$

Moments

$$E(X) = p$$
$$\text{var}(X) = p(1 - p)$$

Addition Rule If $X_1, \ldots, X_k$ are IID Ber(p) random variables, then $X_1 + \cdots + X_k$ is a Bin(k, p) random variable.

Degeneracy If $p = 0$ the distribution is concentrated at 0. If $p = 1$ the distribution is concentrated at 1.

Relation to Other Distributions Ber(p) = Bin(1, p).
6 Binomial Distribution

Abbreviation  Bin\((n,p)\).

Type  Discrete.

Rationale  Sum of \(n\) IID Bernoulli random variables.

Parameters  Real number \(0 \leq p \leq 1\). Integer \(n \geq 1\).

Sample Space  The interval 0, 1, \ldots, \(n\) of the integers.

Probability Mass Function

\[
f(x) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad x = 0, 1, \ldots, n
\]

Moments

\[
E(X) = np
\]
\[
\text{var}(X) = np(1 - p)
\]

Addition Rule  If \(X_1, \ldots, X_k\) are independent random variables, \(X_i\) being Bin\((n_i, p)\) distributed, then \(X_1 + \cdots + X_k\) is a Bin\((n_1 + \cdots + n_k, p)\) random variable.

Normal Approximation  If \(np\) and \(n(1 - p)\) are both large, then

\[
\text{Bin}(n, p) \approx \mathcal{N}(np, np(1 - p))
\]

Poisson Approximation  If \(n\) is large but \(np\) is small, then

\[
\text{Bin}(n, p) \approx \text{Poi}(np)
\]

Theorem  The fact that the probability mass function sums to one is equivalent to the binomial theorem: for any real numbers \(a\) and \(b\)

\[
\sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} = (a + b)^n.
\]
Degeneracy If \( p = 0 \) the distribution is concentrated at 0. If \( p = 1 \) the distribution is concentrated at \( n \).

Relation to Other Distributions \( \text{Ber}(p) = \text{Bin}(1, p) \).

7 Hypergeometric Distribution

Abbreviation Hypergeometric\((A, B, n)\).

Type Discrete.

Rationale Sample of size \( n \) without replacement from finite population of \( B \) zeros and \( A \) ones.

Sample Space The interval \( \max(0, n - B), \ldots, \min(n, A) \) of the integers.

Probability Mass Function

\[
f(x) = \binom{A}{x} \binom{B}{n-x} \binom{A+B}{n}, \quad x = \max(0, n - B), \ldots, \min(n, A)
\]

Moments

\[
E(X) = np
\]
\[
\text{var}(X) = np(1 - p) \cdot \frac{N - n}{N - 1}
\]

where

\[
p = \frac{A}{A + B} \quad (7.1)
\]
\[
N = A + B
\]

Binomial Approximation If \( n \) is small compared to either \( A \) or \( B \), then

\[
\text{Hypergeometric}(n, A, B) \approx \text{Bin}(n, p)
\]

where \( p \) is given by (7.1).
**Normal Approximation**  If $n$ is large, but small compared to either $A$ or $B$, then

$$\text{Hypergeometric}(n, A, B) \approx \mathcal{N}(np, np(1 - p))$$

where $p$ is given by (7.1).

**Theorem**  The fact that the probability mass function sums to one is equivalent to

$$\sum_{x = \max(0, n-B)}^{\min(A,n)} \binom{A}{x} \binom{B}{n-x} = \binom{A+B}{n}$$

---

**8 Poisson Distribution**

**Abbreviation**  Poi($\mu$)

**Type**  Discrete.

**Rationale**  Counts in a Poisson process.

**Parameter**  Real number $\mu > 0$.

**Sample Space**  The non-negative integers 0, 1, . . .

**Probability Mass Function**

$$f(x) = \frac{\mu^x}{x!} e^{-\mu}, \quad x = 0, 1, \ldots$$

**Moments**

$$E(X) = \mu$$

$$\text{var}(X) = \mu$$

**Addition Rule**  If $X_1, \ldots, X_k$ are independent random variables, $X_i$ being Poi($\mu_i$) distributed, then $X_1 + \cdots + X_k$ is a Poi($\mu_1 + \cdots + \mu_k$) random variable.

**Normal Approximation**  If $\mu$ is large, then

$$\text{Poi}(\mu) \approx \mathcal{N}(\mu, \mu)$$
Theorem The fact that the probability mass function sums to one is equivalent to the Maclaurin series for the exponential function: for any real number $x$

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

9 Geometric Distribution

Abbreviation Geo($p$).

Type Discrete.

Rationales

- Discrete lifetime of object that does not age.
- Waiting time or interarrival time in sequence of IID Bernoulli trials.
- Inverse sampling.
- Discrete analog of the exponential distribution.

Parameter Real number $0 < p \leq 1$.

Sample Space The non-negative integers 0, 1, \ldots.

Probability Mass Function

$$
 f(x) = p(1 - p)^x \quad x = 0, 1, \ldots
$$

Moments

$$
 E(X) = \frac{1 - p}{p} \\
 \text{var}(X) = \frac{1 - p}{p^2}
$$

Addition Rule If $X_1, \ldots, X_k$ are IID Geo($p$) random variables, then $X_1 + \cdots + X_k$ is a NegBin($k, p$) random variable.
Theorem  The fact that the probability mass function sums to one is equivalent to the geometric series: for any real number $s$ such that $|s| < 1$

$$\sum_{k=0}^{\infty} s^k = \frac{1}{1-s}.$$ 

Degeneracy  If $p = 1$ the distribution is concentrated at 0.

10 Negative Binomial Distribution

Abbreviation  NegBin$(r,p)$.

Type  Discrete.

Rationale
- Sum of IID geometric random variables.
- Inverse sampling.
- Gamma mixture of Poisson distributions.

Parameters  Real number $0 < p \leq 1$. Integer $r \geq 1$.

Sample Space  The non-negative integers $0, 1, \ldots$.

Probability Mass Function

$$f(x) = \binom{r+x-1}{x} p^r (1-p)^x, \quad x = 0, 1, \ldots$$

Moments

$$E(X) = \frac{r(1-p)}{p}$$
$$\text{var}(X) = \frac{r(1-p)}{p^2}$$

Addition Rule  If $X_1, \ldots, X_k$ are independent random variables, $X_i$ being NegBin$(r_i, p)$ distributed, then $X_1 + \cdots + X_k$ is a NegBin$(r_1 + \cdots + r_k, p)$ random variable.
Normal Approximation  If \( r(1-p) \) is large, then

\[
\text{NegBin}(r,p) \approx \mathcal{N}\left(\frac{r(1-p)}{p}, \frac{r(1-p)}{p^2}\right)
\]

Degeneracy  If \( p = 1 \) the distribution is concentrated at 0.

Extended Definition  The definition makes sense for noninteger \( r \) if binomial coefficients are defined by

\[
\binom{r}{k} = \frac{r \cdot (r-1) \cdots (r-k+1)}{k!}
\]

which for integer \( r \) agrees with the standard definition.

Also

\[
\binom{r+x-1}{x} = (-1)^x \binom{-r}{x} \tag{10.1}
\]

which explains the name “negative binomial.”

Theorem  The fact that the probability mass function sums to one is equivalent to the generalized binomial theorem: for any real number \( s \) such that \(-1 < s < 1\) and any real number \( m \)

\[
\sum_{k=0}^{\infty} \binom{m}{k} s^k = (1 + s)^m. \tag{10.2}
\]

If \( m \) is a nonnegative integer, then \( \binom{m}{k} \) is zero for \( k > m \), and we get the ordinary binomial theorem.

Changing variables from \( m \) to \(-m\) and from \( s \) to \(-s\) and using (10.1) turns (10.2) into

\[
\sum_{k=0}^{\infty} \binom{m+k-1}{k} s^k = \sum_{k=0}^{\infty} \binom{-m}{k} (-s)^k = (1 - s)^{-m}
\]

which has a more obvious relationship to the negative binomial density summing to one.

11 Normal Distribution

Abbreviation  \( \mathcal{N}(\mu, \sigma^2) \).
Type  Continuous.

Rationale
- Limiting distribution in the central limit theorem.
- Error distribution that turns the method of least squares into maximum likelihood estimation.

Parameters  Real numbers $\mu$ and $\sigma^2 > 0$.

Sample Space  The real numbers.

Probability Density Function
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

Moments
$$E(X) = \mu$$
$$\text{var}(X) = \sigma^2$$
$$E\{(X - \mu)^3\} = 0$$
$$E\{(X - \mu)^4\} = 3\sigma^4$$

Linear Transformations  If $X$ is $\mathcal{N}(\mu, \sigma^2)$ distributed, then $aX + b$ is $\mathcal{N}(a\mu + b, a^2\sigma^2)$ distributed.

Addition Rule  If $X_1, \ldots, X_k$ are independent random variables, $X_i$ being $\mathcal{N}(\mu_i, \sigma^2_i)$ distributed, then $X_1 + \cdots + X_k$ is a $\mathcal{N}(\mu_1 + \cdots + \mu_k, \sigma^2_1 + \cdots + \sigma^2_k)$ random variable.

Theorem  The fact that the probability density function integrates to one is equivalent to the integral
$$\int_{-\infty}^{\infty} e^{-z^2/2} \, dz = \sqrt{2\pi}$$

Relation to Other Distributions  If $Z$ is $\mathcal{N}(0, 1)$ distributed, then $Z^2$ is $\text{Gam}(\frac{1}{2}, \frac{1}{2}) = \chi^2(1)$ distributed. Also related to Student $t$, Snedecor $F$, and Cauchy distributions (for which see).
12 Exponential Distribution

Abbreviation  Exp(\(\lambda\)).

Type  Continuous.

Rationales
- Lifetime of object that does not age.
- Waiting time or interarrival time in Poisson process.
- Continuous analog of the geometric distribution.

Parameter  Real number \(\lambda > 0\).

Sample Space  The interval \((0, \infty)\) of the real numbers.

Probability Density Function
\[ f(x) = \lambda e^{-\lambda x}, \quad 0 < x < \infty \]

Cumulative Distribution Function
\[ F(x) = 1 - e^{-\lambda x}, \quad 0 < x < \infty \]

Moments
\[ E(X) = \frac{1}{\lambda}, \quad \text{var}(X) = \frac{1}{\lambda^2} \]

Addition Rule  If \(X_1, \ldots, X_k\) are IID Exp(\(\lambda\)) random variables, then \(X_1 + \cdots + X_k\) is a Gam(\(k, \lambda\)) random variable.

Relation to Other Distributions  Exp(\(\lambda\)) = Gam(1, \(\lambda\)).

13 Gamma Distribution

Abbreviation  Gam(\(\alpha, \lambda\)).
Type  Continuous.

Rationales

• Sum of IID exponential random variables.
• Conjugate prior for exponential, Poisson, or normal precision family.

Parameter  Real numbers $\alpha > 0$ and $\lambda > 0$.

Sample Space  The interval $(0, \infty)$ of the real numbers.

Probability Density Function

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

where $\Gamma(\alpha)$ is defined by (13.1) below.

Moments

$$E(X) = \frac{\alpha}{\lambda}$$
$$\text{var}(X) = \frac{\alpha}{\lambda^2}$$

Addition Rule  If $X_1, \ldots, X_k$ are independent random variables, $X_i$ being $\text{Gam}(\alpha_i, \lambda)$ distributed, then $X_1 + \cdots + X_k$ is a $\text{Gam}(\alpha_1 + \cdots + \alpha_k, \lambda)$ random variable.

Normal Approximation  If $\alpha$ is large, then

$$\text{Gam}(\alpha, \lambda) \approx N\left(\frac{\alpha}{\lambda}, \frac{\alpha}{\lambda^2}\right)$$

Theorem  The fact that the probability density function integrates to one is equivalent to the integral

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

the case $\lambda = 1$ is the definition of the gamma function

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx \quad (13.1)$$
Relation to Other Distributions

- $\text{Exp}(\lambda) = \text{Gam}(1, \lambda)$.
- $\text{ch}^2(\nu) = \text{Gam}(\frac{\nu}{2}, \frac{1}{2})$.
- If $X$ and $Y$ are independent, $X$ is $\Gamma(\alpha_1, \lambda)$ distributed and $Y$ is $\Gamma(\alpha_2, \lambda)$ distributed, then $X/(X+Y)$ is $\text{Beta}(\alpha_1, \alpha_2)$ distributed.
- If $Z$ is $\mathcal{N}(0, 1)$ distributed, then $Z^2$ is $\text{Gam}(\frac{1}{2}, \frac{1}{2})$ distributed.

Facts About Gamma Functions
Integration by parts in (13.1) establishes the gamma function recursion formula

$$\Gamma(\alpha + 1) = \alpha \Gamma(\alpha), \quad \alpha > 0$$  

(13.2)

The relationship between the $\text{Exp}(\lambda)$ and $\text{Gam}(1, \lambda)$ distributions gives

$$\Gamma(1) = 1$$

and the relationship between the $\mathcal{N}(0, 1)$ and $\text{Gam}(\frac{1}{2}, \frac{1}{2})$ distributions gives

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

Together with the recursion (13.2) these give for any positive integer $n$

$$\Gamma(n + 1) = n!$$

and

$$\Gamma(n + \frac{1}{2}) = (n - \frac{1}{2}) \left(n - \frac{3}{2}\right) \cdots \frac{3}{2} \cdot \frac{1}{2} \sqrt{\pi}$$

14 Beta Distribution

Abbreviation  $\text{Beta}(\alpha_1, \alpha_2)$.

Type  Continuous.

Rationales

- Ratio of gamma random variables.
- Conjugate prior for binomial or negative binomial family.
Parameter Real numbers \(\alpha_1 > 0\) and \(\alpha_2 > 0\).

Sample Space The interval \((0, 1)\) of the real numbers.

Probability Density Function
\[
f(x) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} x^{\alpha_1-1}(1-x)^{\alpha_2-1} \quad 0 < x < 1
\]
where \(\Gamma(\alpha)\) is defined by (13.1) above.

Moments
\[
E(X) = \frac{\alpha_1}{\alpha_1 + \alpha_2} \\
\text{var}(X) = \frac{\alpha_1 \alpha_2}{(\alpha_1 + \alpha_2)^2(\alpha_1 + \alpha_2 + 1)}
\]

Theorem The fact that the probability density function integrates to one is equivalent to the integral
\[
\int_0^1 x^{\alpha_1-1}(1-x)^{\alpha_2-1} \, dx = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)}
\]

Relation to Other Distributions

• If \(X\) and \(Y\) are independent, \(X\) is \(\Gamma(\alpha_1, \lambda)\) distributed and \(Y\) is \(\Gamma(\alpha_2, \lambda)\) distributed, then \(X/(X+Y)\) is \(\text{Beta}(\alpha_1, \alpha_2)\) distributed.

• \(\text{Beta}(1, 1) = \text{Unif}(0, 1)\).

15 Multinomial Distribution

Abbreviation \(\text{Multi}(n, p)\).

Type Discrete.

Rationale Multivariate analog of the binomial distribution.
Parameters  Real vector \( \mathbf{p} \) in the parameter space

\[
\left\{ \mathbf{p} \in \mathbb{R}^k : 0 \leq p_i, \ i = 1, \ldots, k, \text{ and } \sum_{i=1}^{k} p_i = 1 \right\}
\]  \hspace{1cm} (15.1)

(real vectors whose components are nonnegative and sum to one).

Sample Space  The set of vectors

\[
S = \left\{ \mathbf{x} \in \mathbb{Z}^k : 0 \leq x_i, \ i = 1, \ldots, k, \text{ and } \sum_{i=1}^{k} x_i = n \right\}
\]  \hspace{1cm} (15.2)

(integer vectors whose components are nonnegative and sum to \( n \)).

Probability Mass Function

\[
f(\mathbf{x}) = \binom{n}{\mathbf{x}} \prod_{i=1}^{k} p_i^{x_i}, \quad \mathbf{x} \in S
\]

where

\[
\binom{n}{\mathbf{x}} = \frac{n!}{\prod_{i=1}^{k} x_i!}
\]

is called a multinomial coefficient.

Moments

\[
E(X_i) = np_i
\]
\[
\text{var}(X_i) = np_i(1 - p_i)
\]
\[
\text{cov}(X_i, X_j) = -np_ip_j, \quad i \neq j
\]

Moments (Vector Form)

\[
E(\mathbf{X}) = n\mathbf{p}
\]
\[
\text{var}(\mathbf{X}) = n\mathbf{M}
\]

where

\[
\mathbf{M} = \mathbf{P} - \mathbf{p}\mathbf{p}^T
\]

where \( \mathbf{P} \) is the diagonal matrix whose vector of diagonal elements is \( \mathbf{p} \).
**Addition Rule**  If $X_1, \ldots, X_k$ are independent random vectors, $X_i$ being $\text{Multi}(n_i, p)$ distributed, then $X_1 + \cdots + X_k$ is a $\text{Multi}(n_1 + \cdots + n_k, p)$ random variable.

**Normal Approximation**  If $n$ is large and $p$ is not near the boundary of the parameter space (15.1), then

$$\text{Multi}(n, p) \approx \mathcal{N}(np, nM)$$

**Theorem**  The fact that the probability mass function sums to one is equivalent to the **multinomial theorem**: for any vector $a$ of real numbers

$$\sum_{x \in S} \binom{n}{x} \prod_{i=1}^{k} a_i^{x_i} = (a_1 + \cdots + a_k)^n$$

**Degeneracy**  If a vector $a$ exists such that $Ma = 0$, then $\text{var}(a^T X) = 0$.

In particular, the vector $u = (1, 1, \ldots, 1)$ always satisfies $Mu = 0$, so $\text{var}(u^T X) = 0$. This is obvious, since $u^T X = \sum_{i=1}^{k} X_i = n$ by definition of the multinomial distribution, and the variance of a constant is zero. This means a multinomial random vector of dimension $k$ is “really” of dimension no more than $k - 1$ because it is concentrated on a hyperplane containing the sample space (15.2).

**Marginal Distributions**  Every univariate marginal is binomial

$$X_i \sim \text{Bin}(n, p_i)$$

Not, strictly speaking marginals, but random vectors formed by collapsing categories are multinomial. If $A_1, \ldots, A_m$ is a partition of the set $\{1, \ldots, k\}$ and

$$Y_j = \sum_{i \in A_j} X_i, \quad j = 1, \ldots, m$$

$$q_j = \sum_{i \in A_j} p_i, \quad j = 1, \ldots, m$$

then the random vector $Y$ has a $\text{Multi}(n, q)$ distribution.
Conditional Distributions  If \( \{i_1, \ldots, i_m\} \) and \( \{i_{m+1}, \ldots, i_k\} \) partition the set \( \{1, \ldots, k\} \), then the conditional distribution of \( X_{i_1}, \ldots, X_{i_m} \) given \( X_{i_{m+1}}, \ldots, X_{i_k} \) is \( \text{Multi}(n - X_{i_{m+1}} - \cdots - X_{i_k}, q) \), where the parameter vector \( q \) has components

\[
q_j = \frac{p_{ij}}{p_{i1} + \cdots + p_{im}}, \quad j = 1, \ldots, m
\]

Relation to Other Distributions

- Each marginal of a multinomial is binomial.
- If \( X \) is \( \text{Bin}(n, p) \), then the vector \( (X, n - X) \) is \( \text{Multi}(n, (p, 1-p)) \).

16  Bivariate Normal Distribution

Abbreviation  See multivariate normal below.

Type  Continuous.

Rationales  See multivariate normal below.

Parameters  Real vector \( \mu \) of dimension 2, real symmetric positive semi-definite matrix \( M \) of dimension \( 2 \times 2 \) having the form

\[
M = \begin{pmatrix}
\sigma_1^2 & \rho \sigma_1 \sigma_2 \\
\rho \sigma_1 \sigma_2 & \sigma_2^2
\end{pmatrix}
\]

where \( \sigma_1 > 0, \sigma_2 > 0 \) and \(-1 < \rho < +1\).

Sample Space  The Euclidean space \( \mathbb{R}^2 \).

Probability Density Function

\[
f(x) = \frac{1}{2\pi} \det(M)^{-1/2} \exp \left( -\frac{1}{2}(x - \mu)^T M^{-1} (x - \mu) \right) = \frac{1}{2\pi \sqrt{1 - \rho^2} \sigma_1 \sigma_2} \exp \left( -\frac{1}{2(1 - \rho^2)} \left( \begin{array}{c}
\frac{x_1 - \mu_1}{\sigma_1} \\
\frac{x_2 - \mu_2}{\sigma_2}
\end{array} \right)^2 - 2\rho \left( \begin{array}{c}
\frac{x_1 - \mu_1}{\sigma_1} \\
\frac{x_2 - \mu_2}{\sigma_2}
\end{array} \right) \left( \begin{array}{c}
\frac{x_1 - \mu_1}{\sigma_1} \\
\frac{x_2 - \mu_2}{\sigma_2}
\end{array} \right)^\top \right), \quad x \in \mathbb{R}^2
\]
Moments

\[ E(X_i) = \mu_i, \quad i = 1, 2 \]
\[ \text{var}(X_i) = \sigma_i^2, \quad i = 1, 2 \]
\[ \text{cov}(X_1, X_2) = \rho \sigma_1 \sigma_2 \]
\[ \text{cor}(X_1, X_2) = \rho \]

Moments (Vector Form)

\[ E(X) = \mu \]
\[ \text{var}(X) = M \]

Linear Transformations  See multivariate normal below.

Addition Rule  See multivariate normal below.

Marginal Distributions  \( X_i \) is \( \mathcal{N}(\mu_i, \sigma_i^2) \) distributed, \( i = 1, 2 \).

Conditional Distributions  The conditional distribution of \( X_2 \) given \( X_1 \) is

\[ \mathcal{N}\left( \mu_2 + \rho \frac{\sigma_2}{\sigma_1} (x_1 - \mu_1), (1 - \rho^2) \sigma_2^2 \right) \]

17 Multivariate Normal Distribution

Abbreviation  \( \mathcal{N}(\mu, M) \)

Type  Continuous.

Rationales

- Multivariate analog of the univariate normal distribution.
- Limiting distribution in the multivariate central limit theorem.

Parameters  Real vector \( \mu \) of dimension \( k \), real symmetric positive semi-definite matrix \( M \) of dimension \( k \times k \).

Sample Space  The Euclidean space \( \mathbb{R}^k \).
Probability Density Function  If $M$ is (strictly) positive definite,

$$f(x) = (2\pi)^{-k/2} \det(M)^{-1/2} \exp\left(-\frac{1}{2}(x - \mu)^T M^{-1} (x - \mu)\right), \quad x \in \mathbb{R}^k$$

Otherwise there is no density ($X$ is concentrated on a hyperplane).

Moments (Vector Form)

$$E(X) = \mu$$
$$\text{var}(X) = M$$

Linear Transformations  If $X$ is $\mathcal{N}(\mu, M)$ distributed, then $a + BX$, where $a$ is a constant vector and $B$ is a constant matrix of dimensions such that the vector addition and matrix multiplication make sense, has the $\mathcal{N}(a + B\mu, BMB^T)$ distribution.

Addition Rule  If $X_1, \ldots, X_k$ are independent random vectors, $X_i$ being $\mathcal{N}(\mu_i, M_i)$ distributed, then $X_1 + \cdots + X_k$ is a $\mathcal{N}(\mu_1 + \cdots + \mu_k, M_1 + \cdots + M_k)$ random variable.

Degeneracy  If a vector $a$ exists such that $Ma = 0$, then $\text{var}(a^T X) = 0$.

Partitioned Vectors and Matrices  The random vector and parameters are written in *partitioned form*

$$X \equiv \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

$$\mu \equiv \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

$$M \equiv \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_2 \end{pmatrix}$$

when $X_1$ consists of the first $r$ elements of $X$ and $X_2$ of the other $k - r$ elements and similarly for $\mu_1$ and $\mu_2$.

Marginal Distributions  Every marginal of a multivariate normal is normal (univariate or multivariate as the case may be). In partitioned form, the (marginal) distribution of $X_1$ is $\mathcal{N}(\mu_1, M_{11})$.  

20
Conditional Distributions  Every conditional of a multivariate normal is normal (univariate or multivariate as the case may be). In partitioned form, the conditional distribution of $X_1$ given $X_2$ is

$$
\mathcal{N}(\mu_1 + M_{12}M_{22}^{-1}(X_2 - \mu_2), M_{11} - M_{12}M_{22}^{-1}M_{21})
$$

where the notation $M_{22}^{-1}$ denotes the inverse of the matrix $M_{22}$ if the matrix is invertible and otherwise any generalized inverse.

18 Chi-Square Distribution

Abbreviation  $\chi^2(\nu)$ or $\chi^2(\nu)$.

Type  Continuous.

Rationales

- Sum of squares of IID standard normal random variables.
- Sampling distribution of sample variance when data are IID normal.
- Asymptotic distribution in Pearson chi-square test.
- Asymptotic distribution of log likelihood ratio.

Parameter  Real number $\nu > 0$ called “degrees of freedom.”

Sample Space  The interval $(0, \infty)$ of the real numbers.

Probability Density Function

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

Moments

$$
E(X) = \nu \\
\text{var}(X) = 2\nu
$$

Addition Rule  If $X_1, \ldots, X_k$ are independent random variables, $X_i$ being $\chi^2(\nu_i)$ distributed, then $X_1 + \cdots + X_k$ is a $\chi^2(\nu_1 + \cdots + \nu_k)$ random variable.
Normal Approximation  If $\nu$ is large, then
\[ \text{chi}^2(\nu) \approx \mathcal{N}(\nu, 2\nu) \]

Relation to Other Distributions

- $\text{chi}^2(\nu) = \text{Gam}(\frac{\nu}{2}, \frac{1}{2})$.
- If $X$ is $\mathcal{N}(0, 1)$ distributed, then $X^2$ is $\text{chi}^2(1)$ distributed.
- If $Z$ and $Y$ are independent, $X$ is $\mathcal{N}(0, 1)$ distributed and $Y$ is $\text{chi}^2(\nu)$ distributed, then $X/\sqrt{Y/\nu}$ is $t(\nu)$ distributed.
- If $X$ and $Y$ are independent and are $\text{chi}^2(\mu)$ and $\text{chi}^2(\nu)$ distributed, respectively, then $(X/\mu)/(Y/\nu)$ is $F(\mu, \nu)$ distributed.

19  Student’s $t$ Distribution

Abbreviation  $t(\nu)$.

Type  Continuous.

Rationales

- Sampling distribution of pivotal quantity $\sqrt{n}(\bar{X}_n - \mu)/S_n$ when data are IID normal.
- Marginal for $\mu$ in conjugate prior family for two-parameter normal data.

Parameter  Real number $\nu > 0$ called “degrees of freedom.”

Sample Space  The real numbers.

Probability Density Function

\[ f(x) = \frac{1}{\sqrt{\nu\pi}} \cdot \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \cdot \frac{1}{\left(1 + \frac{x^2}{\nu}\right)^{(\nu+1)/2}}, \quad -\infty < x < +\infty \]
Moments  If $\nu > 1$, then 
\[ E(X) = 0. \]
Otherwise the mean does not exist. If $\nu > 2$, then 
\[ \text{var}(X) = \frac{\nu}{\nu - 2}. \]
Otherwise the variance does not exist.

Normal Approximation  If $\nu$ is large, then 
\[ t(\nu) \approx N(0, 1) \]

Relation to Other Distributions
- If $X$ and $Y$ are independent, $X$ is $N(0, 1)$ distributed and $Y$ is $\chi^2(\nu)$ distributed, then $X/\sqrt{Y/\nu}$ is $t(\nu)$ distributed.
- If $X$ is $t(\nu)$ distributed, then $X^2$ is $F(1, \nu)$ distributed.
- $t(1) = \text{Cauchy}(0, 1)$.

20 Snedecor’s $F$ Distribution

Abbreviation  $F(\mu, \nu)$.

Type  Continuous.

Rationale
- Ratio of sums of squares for normal data (test statistics in regression and analysis of variance).

Parameters  Real numbers $\mu > 0$ and $\nu > 0$ called “numerator degrees of freedom” and “denominator degrees of freedom,” respectively.

Sample Space  The interval $(0, \infty)$ of the real numbers.

Probability Density Function
\[ f(x) = \frac{\Gamma\left(\frac{\mu+\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{\mu}{2}\right)} \cdot \frac{x^{\mu/2-1}}{(\mu + x)(\mu + \nu)/2}, \quad 0 < x < +\infty \]
Moments  If $\nu > 2$, then

$$E(X) = \frac{\nu}{\nu - 2}.$$  

Otherwise the mean does not exist.

Relation to Other Distributions

- If $X$ and $Y$ are independent and are chi$^2(\mu)$ and chi$^2(\nu)$ distributed, respectively, then $(X/\mu)/(Y/\nu)$ is $F(\mu, \nu)$ distributed.
- If $X$ is $t(\nu)$ distributed, then $X^2$ is $F(1, \nu)$ distributed.

### 21 Cauchy Distribution

**Abbreviation**  Cauchy($\mu, \sigma$).

**Type**  Continuous.

**Rationales**

- Very heavy tailed distribution.
- Counterexample to law of large numbers.

**Parameters**  Real numbers $\mu$ and $\sigma > 0$.

**Sample Space**  The real numbers.

**Probability Density Function**

$$f(x) = \frac{1}{\pi\sigma} \cdot \frac{1}{1 + \left(\frac{x - \mu}{\sigma}\right)^2}, \quad -\infty < x < +\infty$$

**Moments**  No moments exist.

**Addition Rule**  If $X_1, \ldots, X_k$ are IID Cauchy($\mu, \sigma$) random variables, then $\bar{X}_n = (X_1 + \cdots + X_k)/n$ is also Cauchy($\mu, \sigma$).
Relation to Other Distributions

- \( t(1) = \text{Cauchy}(0, 1) \).

## 22 Laplace Distribution

**Abbreviation**  \( \text{Laplace}(\mu, \sigma) \).

**Type**  Continuous.

**Rationales**  The sample median is the maximum likelihood estimate of the location parameter.

**Parameters**  Real numbers \( \mu \) and \( \sigma > 0 \), called the mean and standard deviation, respectively.

**Sample Space**  The real numbers.

**Probability Density Function**

\[
f(x) = \frac{\sqrt{2}}{2\sigma} \exp \left( -\sqrt{2} \left| \frac{x - \mu}{\sigma} \right| \right), \quad -\infty < x < \infty
\]

**Moments**

\[
E(X) = \mu, \quad \text{var}(X) = \sigma^2
\]

## 23 Dirichlet Distribution

**Abbreviation**  none.

**Type**  Continuous. Multivariate.

**Rationales**

- multivariate analog of beta distribution; Dirichlet is to Beta as Multinomial is to Binomial
- Conjugate prior for multinomial family.
Parameter  Vector $\alpha$ having strictly positive real number components.

Sample Space  The “unit simplex,” the set of all probability vectors $p$ having the same dimension as $\alpha$. If $d$ is the dimension of $\alpha$, then the sample space is

$$S = \{ x \in [0,1]^d : x_1 + \cdots + x_d = 1 \}.$$  

Probability Density Function  The Dirichlet distribution is degenerate since the components sum to one with probability one. If we drop the first component, then the remaining components have PDF

$$f(x_2, \ldots, x_d) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_d)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_d)} (1 - x_2 - \cdots - x_d) \prod_{i=2}^d x_i^{\alpha_i-1},$$

$$0 < x_i, i = 2, \ldots, d$$

where the gamma function is defined by (13.1) above.

Moments  Let $I = \{1, \ldots, d\}$. Then

$$E(X_i) = \frac{\alpha_i}{\alpha_1 + \cdots + \alpha_d}$$

$$\text{var}(X_i) = \frac{\alpha_i}{(\alpha_1 + \cdots + \alpha_d)^2(\alpha_1 + \cdots + \alpha_d + 1)} \sum_{j \in I, j \neq i} \alpha_j$$

$$\text{cov}(X_i, X_j) = -\frac{\alpha_i \alpha_j}{(\alpha_1 + \cdots + \alpha_d)^2(\alpha_1 + \cdots + \alpha_d + 1)}$$

Theorem  The fact that the probability density function integrates to one is equivalent to the integral

$$\int \cdots \int (1 - x_2 - \cdots - x_d)^{\alpha_1-1} \left( \prod_{i=2}^d x_i^{\alpha_i-1} \right) dx_2 \cdots dx_d = \frac{\Gamma(\alpha_1) \cdots \Gamma(\alpha_d)}{\Gamma(\alpha_1 + \cdots + \alpha_d)}$$

Marginal Distributions  Every univariate marginal is beta

$$X_i \sim \text{Beta} \left( \alpha_i, \sum_{i \in I, i \neq j} \alpha_j \right)$$
where $I$ is defined under Moments.

Multivariate marginals are “almost” Dirichlet. If we collapse categories, we get another Dirichlet. Let $\mathcal{A}$ be a partition of $I = \{1, \ldots, d\}$, and define

$$Z_A = \sum_{i \in A} Y_i, \quad A \in \mathcal{A}.$$  
$$\beta_A = \sum_{i \in A} \alpha_i, \quad A \in \mathcal{A}.$$  

Then the random vector having components $Z_A$ has the Dirichlet distribution with parameter vector having components $\beta_A$.

**Conditional Distributions**  Conditionals of Dirichlet are constant times Dirichlet. The conditional distribution of $X_1, \ldots, X_k$ given $X_{k+1}, \ldots, X_d$ is determined by the conditional distribution of the random vector $Y$ having components

$$Y_i = \frac{X_i}{1 - X_{k+1} - \cdots - X_d}$$

which has a Dirichlet distribution whose parameter vector has components $\alpha_1, \ldots, \alpha_k$. Multiplying $Y$ by $1 - X_{k+1} - \cdots - X_d$ gives a random vector having the conditional distribution under discussion.

**Relation to Other Distributions**  All univariate marginals are beta. The joint distribution can be written as a function of independent beta random variables.