

# Stat 5101 Notes: Brand Name Distributions

Charles J. Geyer

February 25, 2019

## Contents

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

<b>16 Bivariate Normal Distribution</b>	<b>18</b>
<b>17 Multivariate Normal Distribution</b>	<b>19</b>
<b>18 Chi-Square Distribution</b>	<b>21</b>
<b>19 Student's <math>t</math> Distribution</b>	<b>22</b>
<b>20 Snedecor's <math>F</math> Distribution</b>	<b>23</b>
<b>21 Cauchy Distribution</b>	<b>24</b>
<b>22 Laplace Distribution</b>	<b>25</b>
<b>23 Dirichlet Distribution</b>	<b>25</b>

## **1 Discrete Uniform Distribution**

**Abbreviation**  $\text{DiscUnif}(n)$ .

**Type** Discrete.

**Rationale** Equally likely outcomes.

**Sample Space** The interval  $1, 2, \dots, n$  of the integers.

**Probability Mass Function**

$$f(x) = \frac{1}{n}, \quad x = 1, 2, \dots, n$$

**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**  $\text{Unif}(a, b)$ .

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

### 4 General Uniform Distribution

**Type** Continuous.

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

## Probability Density Function

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

where  $c$  is the measure (length in one dimension, area in two, volume in three, etc.) of the set  $S$ .

## 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

$$\begin{aligned} E(X) &= p \\ \text{var}(X) &= p(1 - p) \end{aligned}$$

**Addition Rule** If  $X_1, \dots, X_k$  are IID Ber( $p$ ) random variables, then  $X_1 + \dots + X_k$  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**  $\text{Bin}(n, p)$ .

**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  $0, 1, \dots, n$  of the integers.

**Probability Mass Function**

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \dots, n$$

**Moments**

$$\begin{aligned} E(X) &= np \\ \text{var}(X) &= np(1-p) \end{aligned}$$

**Addition Rule** If  $X_1, \dots, X_k$  are independent random variables,  $X_i$  being  $\text{Bin}(n_i, p)$  distributed, then  $X_1 + \dots + X_k$  is a  $\text{Bin}(n_1 + \dots + n_k, p)$  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  $\max(0, n - B), \dots, \min(n, A)$  of the integers.

**Probability Mass Function**

$$f(x) = \frac{\binom{A}{x} \binom{B}{n-x}}{\binom{A+B}{n}}, \quad x = \max(0, n - B), \dots, \min(n, A)$$

**Moments**

$$\begin{aligned} E(X) &= np \\ \text{var}(X) &= np(1-p) \cdot \frac{N-n}{N-1} \end{aligned}$$

where

$$\begin{aligned} p &= \frac{A}{A+B} \\ N &= A+B \end{aligned} \tag{7.1}$$

**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**  $\text{Poi}(\mu)$

**Type** Discrete.

**Rationale** Counts in a Poisson process.

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

**Sample Space** The non-negative integers  $0, 1, \dots$

**Probability Mass Function**

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

**Moments**

$$\begin{aligned} E(X) &= \mu \\ \text{var}(X) &= \mu \end{aligned}$$

**Addition Rule** If  $X_1, \dots, X_k$  are independent random variables,  $X_i$  being  $\text{Poi}(\mu_i)$  distributed, then  $X_1 + \dots + X_k$  is a  $\text{Poi}(\mu_1 + \dots + \mu_k)$  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**  $\text{Geo}(p)$ .

**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  $0, 1, \dots$

### Probability Mass Function

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

### Moments

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

**Addition Rule** If  $X_1, \dots, X_k$  are IID  $\text{Geo}(p)$  random variables, then  $X_1 + \dots + X_k$  is a  $\text{NegBin}(k, p)$  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**  $\text{NegBin}(r, p)$ .

**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  $0, 1, \dots$

**Probability Mass Function**

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

**Moments**

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

**Addition Rule** If  $X_1, \dots, X_k$  are independent random variables,  $X_i$  being  $\text{NegBin}(r_i, p)$  distributed, then  $X_1 + \dots + X_k$  is a  $\text{NegBin}(r_1 + \dots + r_k, p)$  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

$$\binom{r+x-1}{x} = (-1)^x \binom{-r}{x} \quad (10.1)$$

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  $m$

$$\sum_{k=0}^{\infty} \binom{m}{k} s^k = (1+s)^m. \quad (10.2)$$

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**

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

**Moments**

$$\begin{aligned} E(X) &= \mu \\ \text{var}(X) &= \sigma^2 \\ E\{(X - \mu)^3\} &= 0 \\ E\{(X - \mu)^4\} &= 3\sigma^4 \end{aligned}$$

**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  $X_1, \dots, X_k$  are independent random variables,  $X_i$  being  $\mathcal{N}(\mu_i, \sigma_i^2)$  distributed, then  $X_1 + \dots + X_k$  is a  $\mathcal{N}(\mu_1 + \dots + \mu_k, \sigma_1^2 + \dots + \sigma_k^2)$  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  $\text{Gam}(\frac{1}{2}, \frac{1}{2}) = \text{chi}^2(1)$  distributed. Also related to Student  $t$ , Snedecor  $F$ , and Cauchy distributions (for which see).

## 12 Exponential Distribution

**Abbreviation**  $\text{Exp}(\lambda)$ .

**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

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

**Addition Rule** If  $X_1, \dots, X_k$  are IID  $\text{Exp}(\lambda)$  random variables, then  $X_1 + \dots + X_k$  is a  $\text{Gam}(k, \lambda)$  random variable.

**Relation to Other Distributions**  $\text{Exp}(\lambda) = \text{Gam}(1, \lambda)$ .

## 13 Gamma Distribution

**Abbreviation**  $\text{Gam}(\alpha, \lambda)$ .

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

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

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

**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*

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx \tag{13.1}$$

### Relation to Other Distributions

- $\text{Exp}(\lambda) = \text{Gam}(1, \lambda)$ .
- $\text{chi}^2(\nu) = \text{Gam}(\frac{\nu}{2}, \frac{1}{2})$ .
- 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**

$$\Gamma(\alpha + 1) = \alpha\Gamma(\alpha), \quad \alpha > 0 \quad (13.2)$$

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

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

## 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**  $\text{Multi}(n, \mathbf{p})$ .

**Type** Discrete.

**Rationale** Multivariate analog of the binomial distribution.

**Parameters** Real vector  $\mathbf{p}$  in the parameter space

$$\left\{ \mathbf{p} \in \mathbb{R}^k : 0 \leq p_i, i = 1, \dots, k, \text{ and } \sum_{i=1}^k p_i = 1 \right\} \quad (15.1)$$

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

**Sample Space** The set of vectors

$$S = \left\{ \mathbf{x} \in \mathbb{Z}^k : 0 \leq x_i, i = 1, \dots, k, \text{ and } \sum_{i=1}^k x_i = n \right\} \quad (15.2)$$

(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**

$$\begin{aligned} E(X_i) &= np_i \\ \text{var}(X_i) &= np_i(1 - p_i) \\ \text{cov}(X_i, X_j) &= -np_i p_j, \quad i \neq j \end{aligned}$$

**Moments (Vector Form)**

$$\begin{aligned} E(\mathbf{X}) &= n\mathbf{p} \\ \text{var}(\mathbf{X}) &= n\mathbf{M} \end{aligned}$$

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  $\mathbf{X}_1, \dots, \mathbf{X}_k$  are independent random vectors,  $\mathbf{X}_i$  being  $\text{Multi}(n_i, \mathbf{p})$  distributed, then  $\mathbf{X}_1 + \dots + \mathbf{X}_k$  is a  $\text{Multi}(n_1 + \dots + n_k, \mathbf{p})$  random variable.

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

$$\text{Multi}(n, \mathbf{p}) \approx \mathcal{N}(n\mathbf{p}, n\mathbf{M})$$

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

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

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

In particular, the vector  $\mathbf{u} = (1, 1, \dots, 1)$  always satisfies  $\mathbf{M}\mathbf{u} = 0$ , so  $\text{var}(\mathbf{u}^T \mathbf{X}) = 0$ . This is obvious, since  $\mathbf{u}^T \mathbf{X} = \sum_{i=1}^k X_i = n$  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  $A_1, \dots, A_m$  is a partition of the set  $\{1, \dots, k\}$  and

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

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

then the random vector  $\mathbf{Y}$  has a  $\text{Multi}(n, \mathbf{q})$  distribution.

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

$$q_j = \frac{p_{i_j}}{p_{i_1} + \dots + p_{i_m}}, \quad j = 1, \dots, m$$

**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  $\boldsymbol{\mu}$  of dimension 2, real symmetric positive semi-definite matrix  $\mathbf{M}$  of dimension  $2 \times 2$  having the form

$$\mathbf{M} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

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

**Sample Space** The Euclidean space  $\mathbb{R}^2$ .

**Probability Density Function**

$$\begin{aligned} f(\mathbf{x}) &= \frac{1}{2\pi} \det(\mathbf{M})^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{M}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) \\ &= \frac{1}{2\pi\sqrt{1 - \rho^2}\sigma_1\sigma_2} \exp\left(-\frac{1}{2(1 - \rho^2)} \left[ \left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 \right. \right. \\ &\quad \left. \left. - 2\rho \left(\frac{x_1 - \mu_1}{\sigma_1}\right) \left(\frac{x_2 - \mu_2}{\sigma_2}\right) + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2 \right] \right), \quad \mathbf{x} \in \mathbb{R}^2 \end{aligned}$$

### Moments

$$\begin{aligned}E(X_i) &= \mu_i, & i = 1, 2 \\ \text{var}(X_i) &= \sigma_i^2, & i = 1, 2 \\ \text{cov}(X_1, X_2) &= \rho\sigma_1\sigma_2 \\ \text{cor}(X_1, X_2) &= \rho\end{aligned}$$

### Moments (Vector Form)

$$\begin{aligned}E(\mathbf{X}) &= \boldsymbol{\mu} \\ \text{var}(\mathbf{X}) &= \mathbf{M}\end{aligned}$$

**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**  $\mathcal{N}(\boldsymbol{\mu}, \mathbf{M})$

**Type** Continuous.

### Rationales

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

**Parameters** Real vector  $\boldsymbol{\mu}$  of dimension  $k$ , real symmetric positive semi-definite matrix  $\mathbf{M}$  of dimension  $k \times k$ .

**Sample Space** The Euclidean space  $\mathbb{R}^k$ .

**Probability Density Function** If  $\mathbf{M}$  is (strictly) positive definite,

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

Otherwise there is no density ( $\mathbf{X}$  is concentrated on a hyperplane).

**Moments (Vector Form)**

$$\begin{aligned} E(\mathbf{X}) &= \boldsymbol{\mu} \\ \text{var}(\mathbf{X}) &= \mathbf{M} \end{aligned}$$

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

**Addition Rule** If  $\mathbf{X}_1, \dots, \mathbf{X}_k$  are independent random vectors,  $\mathbf{X}_i$  being  $\mathcal{N}(\boldsymbol{\mu}_i, \mathbf{M}_i)$  distributed, then  $\mathbf{X}_1 + \dots + \mathbf{X}_k$  is a  $\mathcal{N}(\boldsymbol{\mu}_1 + \dots + \boldsymbol{\mu}_k, \mathbf{M}_1 + \dots + \mathbf{M}_k)$  random variable.

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

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

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \tag{17.1a}$$

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \tag{17.1b}$$

$$\mathbf{M} = \begin{pmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_2 \end{pmatrix} \tag{17.1c}$$

when  $\mathbf{X}_1$  consists of the first  $r$  elements of  $\mathbf{X}$  and  $\mathbf{X}_2$  of the other  $k - r$  elements and similarly for  $\boldsymbol{\mu}_1$  and  $\boldsymbol{\mu}_2$ .

**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  $\mathbf{X}_1$  is  $\mathcal{N}(\boldsymbol{\mu}_1, \mathbf{M}_{11})$ .

**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  $\mathbf{X}_1$  given  $\mathbf{X}_2$  is

$$\mathcal{N}(\boldsymbol{\mu}_1 + \mathbf{M}_{12}\mathbf{M}_{22}^{-1}[\mathbf{X}_2 - \boldsymbol{\mu}_2], \mathbf{M}_{11} - \mathbf{M}_{12}\mathbf{M}_{22}^{-1}\mathbf{M}_{21})$$

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

## 18 Chi-Square Distribution

**Abbreviation**  $\text{chi}^2(\nu)$  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

$$\begin{aligned} E(X) &= \nu \\ \text{var}(X) &= 2\nu \end{aligned}$$

**Addition Rule** If  $X_1, \dots, X_k$  are independent random variables,  $X_i$  being  $\text{chi}^2(\nu_i)$  distributed, then  $X_1 + \dots + X_k$  is a  $\text{chi}^2(\nu_1 + \dots + \nu_k)$  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**

$$f(x) = \frac{1}{\sqrt{\nu\pi}} \cdot \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \cdot \frac{1}{\left(1 + \frac{x^2}{\nu}\right)^{(\nu+1)/2}}, \quad -\infty < x < +\infty$$

**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

$$t(\nu) \approx \mathcal{N}(0, 1)$$

### Relation to Other Distributions

- If  $X$  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  $X$  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

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

**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  $\text{chi}^2(\mu)$  and  $\text{chi}^2(\nu)$  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  $X_1, \dots, X_k$  are IID Cauchy( $\mu, \sigma$ ) random variables, then  $\bar{X}_n = (X_1 + \dots + X_k)/n$  is also Cauchy( $\mu, \sigma$ ).



### Relation to Other Distributions

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

## 22 Laplace Distribution

**Abbreviation** 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

$$\begin{aligned} E(X) &= \mu \\ \text{var}(X) &= \sigma^2 \end{aligned}$$

## 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  $\mathbf{p}$  having the same dimension as  $\alpha$ . If  $d$  is the dimension of  $\alpha$ , then the sample space is

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

**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

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

$$0 < x_i, i = 2, \dots, d \text{ and } x_2 + \cdots + x_d < 1$$

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

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

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

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

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

**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

$$X_i \sim \text{Beta} \left( \alpha_i, \sum_{\substack{j \in I \\ j \neq i}} \alpha_j \right)$$

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  $I = \{1, \dots, d\}$ , 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  $X_1, \dots, X_k$  given  $X_{k+1}, \dots, X_d$  is determined by the conditional distribution of the random vector  $\mathbf{Y}$  having components

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

which has a Dirichlet distribution whose parameter vector has components  $\alpha_1, \dots, \alpha_k$ . Multiplying  $\mathbf{Y}$  by  $1 - X_{k+1} - \dots - X_d$  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.