Econometric Analysis
Review on Probability and Distribution Theory

by

Jin-Lung Lin

Fall, 2007

1 Two simple questions

1. Toll a die and find the probability that the number is 3.
2. Randomly pick a number, x, between [0,1] and find the probability that P(x=0.1)

Dilemma: infinitely many numbers between 0 and 1.
Should define $P(x = 0.1) = 0$ or $P(x = 0.1) > 0$?

2 Why do we care for mathematical statistics?

- Randomness is the heart of statistics:
  lottery, stock price, effect of a new drug, result of world series, presidential election result, trajectory of a missile, grade of this course, etc.
- Statistics is about inference: from sample to population,
- A solid understanding of statistical theory is important for both math/stat major and for applications in various fields
- Mathematical statistics is bridge to the sophisticated statistical theory.

3 Probability space

Some terminology:
sample space, simple event (sample point), sample space, discrete sample space, event, probability
Probability measure:

1. Axiom 1: $P(A) \geq 0$;

2. Axiom 2: $P(\Omega) = 1$;

3. Axiom 3: If $A_1, A_2, A_3, \cdots$ form a sequence of pairwise mutually exclusive events in Ω ($A_i \cap A_j = 0$ if $i \neq j$), then $P(A_1 \cup A_2 \cup A_3 \cup \cdots) = \sum_{i=1}^{\infty} P(A_i)$

Conditional probability and independence of events
conditional probability vs unconditional probability:
partial information
example: randomly draw a card and find the probability that the card is heart A.
Given that information that the card is ”red”, find the probability that the card is hear A.

Conditional probability of an event A given event B is equal to

$$P(A|B) = \frac{P(A \cap B)}{A(B)}$$

provided $P(B) > 0$.
The multiplicative law of probability:

$$\begin{aligned} P(A \cap B) &= P(A)P(B|A) \\ &= P(B)P(A|B) \end{aligned}$$

If A and B are independent,

$$\begin{aligned} P(A \cap B) &= P(A)P(B) \\ P(A|B) &= P(A) \\ P(B|A) &= P(B) \end{aligned}$$

$$\begin{aligned} P(A \cap B \cap C) &= P[(A \cap B) \cap C] \\ &= P(A \cap B)P(C|A \cap B) \\ &= P(A)P(B|A)P(C|A \cap B) \end{aligned}$$

More generally,

$$P(A_1 \cap A_2 \cap, \cdots, A_k) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_k|A_1 \cap A_2 \cap, \cdots, A_{k-1})$$

If $A_1, \cdots, A_k$ are mutually independent, then

$$P(A_1 \cap A_2 \cap, \cdots, A_k) = P(A_1)P(A_2)P(A_3) \cdots P(A_k)$$

4 Random sampling

sampling with and without replacement
sampling scheme

5 Random variable

A random variable is a real-value function defined over a sample space.

1. a real-value function
2. over a sample space

A random variable is discrete if it can only take finite or countably many number of distinct values

6 Probability distribution of discrete random variables

Notation: $P(Y = y)$
Y : denotes random variable, y denotes particular values, $P(Y = y)$ denotes the probability that Y takes on the value of y, defined as the sum of probability of all sample points in S that are assigned to the value of y. Sometimes, it is denoted as $p(y)$, the probability function (distribution) for Y.

Example: Pick two persons randomly from a pool of 3 men and 3 women, and Y is the numbers of women selected. Answer:

$$p(y) = \frac{\binom{3}{y}\binom{3}{2-y}}{\binom{6}{2}}$$

Theorem 3.1
For any discrete probability distribution, the following must be true:

1. $0 \leq p(y) \leq 1 \quad \forall y$
2. $\sum_y p(y) = 1$

7 Expected values

Definition: $E(Y) = \sum_y yp(y) = \mu$
absolute convergence:
Remark: Expected value exists if it is absolutely convergent, ie. $\sum_y |y|p(y) < \infty$
Theorem 3.2
Let Y is a discrete random variable with probability distribution $p(y)$, and $g(Y)$ be a real-valued function of Y. Then, $E[g(Y)] = \sum_y g(y)p(y)$

The variance of a random variable Y is defined as expected value of $(Y - \mu)^2$, ie $V(Y) = E(Y - \mu)^2$. Standard deviation of Y is $\sqrt{V(Y)}$.

Example
Theorem 3.3
Let Y be a discrete random variable with probability function $p(y)$ and let c be a constant. Then $E(c) = c$.
Theorem 3.4
Let Y be a discrete random variable with probability function $p(y), g(Y)$ be a function of Y, and let c be a constant. Then $E[cg(Y)] = cE[g(Y)]$
Theorem 3.5
Let Y be a discrete random variable with probability function $p(y)$ and $g_1(Y), g_2(Y), \cdots, g_k(Y)$ be k functions of Y. Then, $E[g_1(Y)+g_2(Y)+\cdots+g_k(Y)] = E[g_1(Y)]+E[g_2(Y)]+\cdots+E[g_k(Y)]$
Theorem 3.6
Let Y be a discrete random variable with probability function $p(y)$, then

$$V(Y) = E(Y - \mu)^2 = E(Y^2) - \mu^2$$

8 Binomial probability distribution

identical and independent trials
a binomial experiments has the following properties:

1. consists of fixed number, n of identical trials
2. two outcomes in each trial, S(success) or F(failure).
3. $p(S) = p, p(F) = 1 - p(S) = 1 - p = q$
4. independent trials

5. number of S during the n trials is of interest

Definition: A random variable has binomial distribution based upon n trials with success probability p iff

$$p(y) = \begin{pmatrix} n \\ y \end{pmatrix} p^y q^{(n-y)}, \ \ y = 0, 1, \cdots, n \text{ and } 0 \leq p \leq 1$$

binomial expansion

$$(q+p)^n = \begin{pmatrix} n \\ 0 \end{pmatrix} q^n + \begin{pmatrix} n \\ 1 \end{pmatrix} p^1 q^{(n-1)} + \begin{pmatrix} n \\ 2 \end{pmatrix} p^2 q^{n-2} + \cdots + \begin{pmatrix} n \\ n \end{pmatrix} p^n$$

$\sum_y p(y) = \sum_{y=0}^n \begin{pmatrix} n \\ y \end{pmatrix} p^y q^{(n-y)} = (p+q)^n = 1$

Theorem 3.7

Let Y be the binomial random variable based upon n trials and success probability p. then, $\mu = E(Y) = np$ and $\sigma^2 = V(Y) = npq$

9 Poisson distribution

$P($no accident occur in a subinterval$) = 1 - p$

$P($one accident occur in a subinterval$) = p$

$P($more than one accident occur in a subinterval$) = 0$

Let $\lambda = np$, take the limit of binomial distribution

$$p(y) = \begin{pmatrix} n \\ y \end{pmatrix} p^y (1-p)^{n-y} \rightarrow \frac{\lambda^y}{y!} e^{\lambda} \text{ as } n \rightarrow \infty$$

Definition: A random variable Y has a Poisson distribution iff

$$p(y) = \frac{\lambda^y}{y!} e^{\lambda} \, y = 0, 1, 2, \cdots, \quad \lambda > 0$$

Theorem 3.11

If Y has Poisson distribution, then $\mu = E(Y) = \lambda, \quad \sigma^2 = V(Y) = \lambda$

10 Continuous random variables

A continuous random variable takes on any value in an interval

Theorem 4.1: Properties of a distribution function. If $F(y)$ is a distribution function, then

1. $F(-\infty) \equiv \lim_{y\to-\infty} F(y) = 0$.
2. $F(\infty) \equiv \lim_{y\to\infty} F(y) = 1$.
3. $F(y)$ is a nondecreasing function of y, and is right continuous.

Definition: Let Y be a random variable with distribution $F(y).Y$ is continuous if $F(y)$ is continuous for $-\infty < y < \infty$.
Definition: Let $F(y)$ be the distribution function for a continuous random variable Y. Then, $f(y)$ given by

$$f(y) = \frac{dF(y)}{dy} = F'(y)$$

wherever the derivative exists, is called the probability density function for the random variable Y.
Theorem 4.2: Properties of a density function. If $f(y)$ is a density function, then

1. $f(y) \geq 0 \quad \forall y$
2. $\int_{-\infty}^{\infty} f(y)dy = 1$

Theorem 4.3 If the random variable Y has the density function $f(y)$ and $a \leq b$, then

$$P(a \leq Y \leq b) = \int_a^b f(y)dy$$

11 Expected values for continuous random variables

Definition: The expected value of a random variable Y is

$$E(Y) = \int_{-\infty}^{\infty} yf(y)dy$$

provided that the integral exists ($\int_{-\infty}^{\infty} |y|f(y)dy \leq \infty$).
Theorem 4.4: Let $g(Y)$ be a function of Y. then the expected value of $g(Y)$ is given by

$$E[g(Y)] = \int_{-\infty}^{\infty} g(y)f(y)dy$$

provided that the integral exists.
Theorem 4.5: Let c be a constant, $g(Y), g_1(Y), g_2(Y), \cdots, g_k(Y)$ be $k+1$ functions of a continuous random variable Y. Then,

1. $E(c) = c$.
2. $E[cg(Y)] = cE[(g(Y)]$.
3. $E[g_1(Y) + g_2(Y) + \cdots + g_k(Y)] = E[g_1(Y)] + E[g_2(Y)] + \cdots + E[g_k(Y)]$

12 Uniform probability distribution

Definition: If $\theta_1 < \theta_2$, a random variable Y has a continuous uniform probability distribution on the interval (θ_1, θ_2) iff the density function of Y is:

$$f(y) = \begin{cases} \frac{1}{\theta_1 - \theta_2} & \theta_1 \le y \le \theta_2 \\ 0 & \text{otherwise} \end{cases}$$

Theorem 4.6: If $\theta_1 < \theta_2$ and Y is a random variable uniformly distributed on the interval (θ_1, θ_2), then

$$\mu = E(Y) = \frac{\theta_1 + \theta_2}{2}, \quad \text{and} \quad \sigma^2 = V(Y) = \frac{(\theta_2 - \theta_1)^2}{12}$$

13 Normal probability distribution

Definition: A random variable Y has a normal probability distribution function iff for $\sigma > 0$ and $-\infty < y < \infty$, the density function of Y is:

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

Theorem 4.7 If Y is normally distributed with parameter μ and σ. Then,

$$E(Y) = \mu, \quad V(Y) = \sigma^2$$

14 Bivariate and multivariate probability distributions

Definition Let Y_1, Y_2 be discrete-random variables. The joint probability (bivariate) distribution for Y_1, Y_2 is given by :

$$p(y_1, y_2) = P(Y_1 = y_1, Y_2 = y_2), \quad -\infty < y_1 < \infty, -\infty < y_2 < \infty$$

Theorem 5.1: If Y_1, Y_2 are discrete random variables with joint probability function $p(y_1, y_2)$ then

1. $p(y_1, y_2) \ge 0$ for all y_1, y_2.
2. $\sum_{y_1, y_2} p(y_1, y_2) = 1$ where the sum is over all values (y_1, y_2) that are assigned nonzero probabilities. Properties of a distribution function.

Definition For any random variables Y_1, Y_2, the joint (bivariate) distribution $F(y_1, y_2)$ is given by

$$F(y_1, y_2) = P(Y_1 \leq y_1, Y_2 \leq y_2), \quad -\infty < y_1 < \infty, -\infty < y_2 < \infty$$

Definition Let Y_1, Y_2 be continuous random variables with joint distribution function $F(y_1, y_2)$. The there exist a nonnegative function $f(y_1, y_2)$ such that

$$F(y_1, y_2) = \int_{-\infty, y_1}^{y_1} \int_{-\infty}^{y_2} f(t_1, t_2) dt_1, dt_2 \text{ for all } -\infty < y_1 < \infty, -\infty < y_2 < \infty$$

then, Y_1, Y_2 are said to be jointly continuous random variables with joint density function $f(y_1, y_2)$.
Theorem 5.2: If Y_1, Y_2 are random variables with joint probability function $F(y_1, y_2)$ then

1. $F(-\infty, -\infty) = F(-\infty, y_2) = F(y_1, -\infty) = 0.$
2. $F(\infty, \infty) = 1$
3. if $y_1^* \geq y_1$ and $y_2^* \geq y_2$, then

$$F(y_1^*, y_2^*) - F(y_1^*, y_2) - F(y_1, y_2^*) + F(y_1, y_2) \geq 0$$

.

Theorem 5.3: If Y_1, Y_2 are joint random variables with joint density function $f(y_1, y_2)$ then

1. $f(y_1, y_2) \geq 0$ for all y_1, y_2.
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(y_1, y_2) dy_1 dy_2 = 1$

15 marginal and conditional probability distributions

Definition 5.4 a Let Y_1 and Y_2 be jointly discrete random variables with probability function $p(y_1, y_2)$. Then the marginal probability functions of Y_1 and Y_2, respectively, are given by

$$p_1(y_1) = \sum_{y_2} p(y_1, y_2) \text{ and } p_2(y_2) = \sum_{y_1} p(y_1, y_2).$$

b Let Y_1 and Y_2 be jointly continuous random variables with density function $f(y_1, y_2)$. Then the marginal probability functions of Y_1 and Y_2, respectively, are given by

$$f_1(y_1) = \int_{-\infty}^{\infty} f(y_1, y_2) dy_2 \text{ and } f_2(y_2) = \int_{-\infty}^{\infty} f(y_1, y_2) dy_1.$$

Definition 5.5 If Y_1 and Y_2 are jointly discrete random variables with joint probability function $p(y_1, y_2)$ and marginal probability functions $p(y_1)$ and $p(y_2)$, respectively, then the conditional discrete probability function of Y_1 given Y_2 is

$$p(y_1|y_2) = P(Y_1 = y_1|Y_2 = y_2) = \frac{P(Y_1 = y_1, Y_2 = y_2)}{P(Y_2 = y_2)} = \frac{p(y_1, y_2)}{p_2(y_2)}$$

provided that $p_2(y_2) > 0$.

Definition 5.6 If Y_1 and Y_2 are jointly continuous random variables with joint density function $f(y_1, y_2)$, then the conditional distribution function of Y_1 given $Y_2 = y_2$ is

$$F(y_1|y_2) = P(Y_1 \leq y_1|Y_2 = y_2)$$

Definition 5.7 Let Y_1 and Y_2 be jointly continuous random variables with joint density $f(y_1, y_2)$ and marginal densities $f_1(y_1)$ and $f_2(y_2)$, respectively. For any y_2 such that $f_2(y_2) > 0$, the conditional density of Y_1 given y_2 is given by

$$f(y_1|y_2) = \frac{f(y_1, y_2)}{f_2(y_2)}$$

and, for any y_1 such that $f_1 > 0$, the conditional density of Y_2 given $Y_1 = y_1$ is given by

$$f(y_2|y_1) = \frac{f(y_1, y_2)}{f_1(y_1)}$$

16 Independent random variable

Definition 5.8 Let Y_1 have distribution function $F_1(y_1)$, Y_2 have distribution function $F_2(y_2)$, and Y_1 and Y_2 have joint distribution function $F(y_1, y_2)$. Then Y_1 and Y_2 are said to be independent if and only if

$$F(y_1, y_2) = F_1(y_1)F_2(y_2)$$

for every pair of real numbers (y_1, y_2). If Y_1 and Y_2 are not independent, they are said to be dependent.

Theorem 5.4 If Y_1 and Y_2 are discrete random variables with joint probability function $p(y_1, y_2)$ and marginal probability functions $p_1(y_1)$ and $p_2(y_2)$, respectively, then Y_1 and Y_2 are the independent iff

$$p(y_1, y_2) = p_1(y_1)p_2(y_2)$$

for all pairs of real numbers (y_1, y_2).
If Y_1 and Y_2 are continuous random variables with joint density function $f(y_1, y_2)$ and marginal density functions $f_1(y_1)$ and $f_2(y_2)$, respectively, then Y_1 and Y_2 are independent if and only if

$$f(y_1, y_2) = f_1(y_1)f_2(y_2)$$

17 Expected value of function of random variables

Definition 5.9 Let $g(Y_1, Y_2, \ldots, Y_k)$ be a function of the discrete random variables, $Y_1, Y_2, \ldots, Y_k$, which have probability function $p(y_1, y_2, \ldots, y_k)$. Then the expected value of $g(Y_1, Y_2, \ldots, Y_k)$ is

$$E[g(Y_1, Y_2, \ldots, Y_k)] = \sum_{yk} \cdots \sum_{y2} \sum_{y1} g(y_1, y_2, \ldots, y_k)p(y_1, y_2, \ldots, y_k).$$

If $Y_1, Y_2, \ldots, Y_k$ are continuous random variables with joint density function $f(y_1, y_2, \ldots, f_k)$, then

$$E[g(Y_1, Y_2, \ldots, Y_k] = \int_{yk} \cdots \int_{y2} \int_{y1} g(y_1, y_2, \ldots, y_k) \times f(y_1, y_2, \ldots, y_k) dy_1 dy_2 \ldots dy_k.$$

Theorem 5.6 Let c be a constant. Then

$$E(c) = c. \tag{1}$$

Theorem 5.7 Let $g(Y_1, Y_2)$ be a function of the random variables Y_1, Y_2, and let c be a constant. Then

$$E[cg(Y_1, Y_2] = cE[g(Y_1, Y_2].$$

Theorem 5.8 Let Y_1 and Y_2 be random variables and $g_1(Y_1, Y_2), vg_2(Y_1, Y_2), v \ldots, vg_k(Y_1$ be functions of Y_1 and Y_2. Then

$$\begin{aligned} &E[g_1(Y_1, Y_2) + g_2(Y_1, Y_2) + \ldots + g_k(Y_1, Y_2)] \\ = &E[g_1(Y_1, Y_2)] + E[g_2(Y_1, Y_2)] + \ldots + E[g_k(Y_1, Y_2]. \end{aligned}$$

Theorem 5.9 Let Y_1 and Y_2 be independent random variables and $g(Y_1)$ and $g(Y_2)$ be functions of only Y_1 and Y_2, respectively. Then

$$E[g(Y_1)h(Y_2)] = E[g(Y_1)]E[h(Y_2)].$$

provided that the expectations exist.

18 Covariance of two random variables

Definition 5.10 Let Y_1 and Y_2 be random variables with means μ_1 and μ_2, respectively, the covariance of Y_1 and Y_2 is given by

$$\text{Cov}(Y_1, Y_2) = E[(Y_1 - \mu_1)(Y_2 - \mu_2)].$$

Theorem 5.10 Let Y_1 and Y_2 be random variables with means μ_1 and μ_2, respectively, then

$$\text{Cov}(Y_1, Y_2) = E[(Y_1 - \mu_1)(Y_2) - \mu_2)] = E(Y_1 Y_2) - E(Y_1)E(Y_2).$$

Theorem 5.11 Let Y_1 and Y_2 be independent random variables, then

$$\text{Cov}(Y_1, Y_2) = 0.$$

Theorem 5.12 Let $Y_1, Y_2, \ldots, Y_k$ and $X_1, X_2, \ldots, X_k$ be random variables with $E(Y_i) = \mu_i$ and $E(X_i) = \xi_j$. Define

$$\begin{aligned} U_1 &= \sum_{i=1}^{n} a_i Y_i \quad \text{and} \\ U_2 &= \sum_{i=1}^{n} b_i X_i \end{aligned}$$

for constants $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_m$. Then the following hold:

1. $E(U_1) = \sum_{i=1}^{n} a_i \mu_i$.
2. $V(U_1) = \sum_{i=1}^{n} a_i^2 V(Y_i) + 2 \sum\sum_{i<j} a_i a_j \text{Cov}(Y_i, Y_j)$, where the double sum is over all pairs (i, j) with $i < j$.
3. $\text{Cov}(U_1, U_2) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j \text{Cov}(Y_i, X_j)$.

19 Conditional expectation

Definition 5.13 If Y_1 and Y_2 are any two random variables, the condition expectation of $g(Y_1)$, given that $Y_2 = y_2$, is defined to be

$$E(g(Y_1)|Y_2 = y_2) = \int_{-\infty}^{\infty} g(y_1) f(y_1|y_2) dy_1$$

if Y_1 and Y_2 are jointly continuous and

$$E(g(Y_1)|Y_2 = y_2) = \sum_{y_1} g(y_1) p(y_1|y_2)$$

if Y_1 and Y_2 are jointly discrete.

A random variable may always be written as

$$
\begin{aligned}
y &= E(y|x) + (y - E(y|x)) \\
&= E(Y|x) + \epsilon
\end{aligned}
$$

If $E(y|x) = \alpha + \beta x$, then

$$
\begin{aligned}
\alpha &= E(y) - \beta E(x) \\
\beta &= \frac{cov(x, y)}{Var(x)}
\end{aligned}
$$

Theorem 5.14 Let Y_1 and Y_2 denote random variables. Then

$$E(Y_1) = E[E(Y_1|Y_2)]$$

where, on the right-hand side, the inside expectation is with respect to the conditional distribution of Y_1 given Y_2, and the outside expectation is with respect to the distribution of Y_2.

Theorem 5.15 Let Y_1 and Y_2 denote random variables. Then

$$
\begin{aligned}
V(Y_1) &= E[V(Y_1|Y_2)] + V[E(Y_1|Y_2)] \\
E_{Y_1}(Var(Y_1|Y_2)) &= Var(Y_1) - Var_{Y_1}(E(Y_1|Y_2))
\end{aligned}
$$

20 Bivariate normal distributions

$$
\begin{aligned}
f(y_1, y_2) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-1/2Q} \\
Q &= \frac{1}{1-\rho^2}[\frac{(y_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(y_1-\mu_1)(y_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}]
\end{aligned}
$$

$$
\begin{aligned}
f_1(y_1) &= N(\mu_1, \sigma_1^2) \\
f_2(y_2) &= N(\mu_2, \sigma_2^2) \\
f(y_2|y_1) &= N(\alpha + \beta y_1, \sigma_2^2(1-\rho^2))
\end{aligned}
$$

where $\alpha = \mu_2 - \beta\mu_1, \beta = \frac{\sigma_{12}}{\sigma_1^2}$

21 Finding the probability distribution of a function of random variable

Four methods to find the distribution $U = f(Y_1, Y_2, \ldots, Y_n)$

1. method of distribution functions
2. method of transformations
3. method of moment-generating functions
4. order statistics

22 Some frequently used distributions

1. χ^2 distribution

   - If $z \sim N(0,1)$, then $x = z^2 \sim \chi^2(1)$ that is chi-squared with one degree of freedom, denoted as
   $$x \sim \chi^2(1)$$
   - If $x_1, \cdots, x_n$ are n independent $\chi^2(1)$, then
   $$\sum_{i=1}^{n} x_i \sim \chi^2(n)$$
   - If $z_i, i = 1, \cdots, n$ are independent $N(0, \sigma^2)$, then
   $$\sum_{i=1}^{n} (z_i/\sigma)^2 \sim \chi^2(n)$$
   - If x_1, x_2 are independent $\chi^2(n_1), \chi^2(n_2)$, then
   $$x_1 + x_2 \sim \chi^2(n_1 + n_2)$$
   - For $v > 30$, $\chi^2(v)$ can be well approximated by
   $$\begin{aligned} z &= (2x)^{1/2} - (2n-1)^{1/2} \\ Prob(\chi^2(n) \leq a) &\approx \Phi[(2a)^{1/2} - (2n-1)^{1/2}] \end{aligned}$$

2. t-distributions
   If $z \sim N(0,1)$ is independent of $x \sim \chi^2(n)$ then
   $$t(n) = \frac{z}{\sqrt{x/n}}$$
   has $t-distribution$ with n degree of freedom

3. F-distributions
   If x_1, x_2 are independent $\chi^2(n_1), \chi^2(n_2)$, then
   $$\frac{x_1/n_1}{x_2/n_2} \sim F(n_1, n_2)$$

4. If $t \sim t(n)$, then $t^2 \sim F(1,n)$

5. For $F(n_1, n_2)$ when n_2 becomes infinite, $\chi(n_2)/n_2$ converges identically to one, $x = n_1 F$ is approximately $\chi^2(n_1)$

6. If $z \sim N(\mu, \Sigma)$ is multivariate normally distributed with mean μ and covariance Σ, then $(z-\mu)'\Sigma^{-1}(z-\mu) \sim \chi^2(J)$ where J is the dimension of z.

7. Lognormal distributions
   $X \sim LN(\mu, \sigma^2)$ if
   $$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-1/2[(lnx-\mu)/\sigma]^2}, x > 0$$
   $$\begin{aligned} E(x) &= e^{\mu+\sigma^2/2} \\ Var(x) &= e^{2\mu+\sigma^2}(e^{\sigma^2}-1) \end{aligned}$$
   If $y \sim LN(\mu, \sigma^2)$, then $lny \sim N(\mu, \sigma^2)$

8. Gamma, beta, and exponential distributions

   - gamma: $f(x) = \frac{\lambda^P}{\Gamma(P)} e^{-\lambda x} x^{P-1}, \quad x \geq 0, \lambda > 0, P > 0$
   - inverse gamma:
     x has the gamma distribution, then $1/x$ is inverse-gamma distributed.
     $f(y) = \frac{\lambda^P}{\Gamma(P)} e^{-\lambda/y} y^{-(P+1)}, y \geq 0, \lambda > 0, P > 0$
   - beta:
     $$f(x0 = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}(\frac{x}{c})^{\beta-1}\frac{1}{c}$$
     where $c > 0$