Probability and Statistics I

STAT 3600 – Fall 2021

Le Chen
lzc0090@auburn.edu

Last updated on

July 4, 2021

Auburn University
Auburn AL

Chapter 5. Distributions of Functions of Random Variables

Let X be a random variables of the continuous type. If we consider a function of X, say $Y = u(X)$, then Y must also be a random variable that has its own distribution. If we can find its distribution function, say,

$$G(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(u(X) \leq y),$$

then its pdf is given by $g(y) = G'(y)$.

Example 5.1-1 Let X have a gamma distribution with pdf

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

where $\alpha > 0, \theta > 0$. Let $Y = e^X$, so that the support of Y is $1 < y < \infty$. For each y in the support, the distribution function of Y is

$$G(y) = \mathbb{P}(Y \le y) = \mathbb{P}(e^X \le y) = \mathbb{P}(X \le \ln y).$$

That is,

$$G(y) = \int_0^{\ln y} \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-x/\theta} dx$$

and thus the pdf $g(y) = G'(y)$ of Y is

$$g(y) = \frac{1}{\Gamma(\alpha)\theta^\alpha} (\ln y)^{\alpha-1} e^{-(\ln y)/\theta} \left(\frac{1}{y}\right), \qquad 1 < y < \infty.$$

Equivalently, we have

$$g(y) = \frac{1}{\Gamma(\alpha)\theta^\alpha} \frac{(\ln y)^{\alpha-1}}{y^{1+1/\theta}} \qquad 1 < y < \infty.$$

which is called a *loggamma* pdf.

Determination of $g(y)$ from $f(x)$

Let X be a continuous r.v. with pdf $f(x)$. If the transformation $y = u(x)$ is one-to-one and has the inverse transformation

$$x = u^{-1}(y) = v(y)$$

then the pdf of Y is given by

$$g(y) = f(v(y))\,|v'(y)|, \quad y \in S_y,$$

where S_y is the support of Y.

Example 5.1-2 Let $Y = 2X + 3$. Find the pdf of Y if X is a uniform r.v. over $(-1, 2)$.

Ans: $g(y) = \begin{cases} \frac{1}{6}, & \text{if } 1 < y < 7, \\ 0, & \text{otherwise.} \end{cases}$

Example 5.1-3 Let X have the pdf $f(x) = xe^{-x^2/2}, 0 < x < \infty$. Find the pdf of $Y = X^2$.
Ans: $g(y) = \cdots$

Here are some examples when the transformation $Y = u(X)$ is not one-to-one.

Example 5.1-4 Let $Y = X^2$. Find the pdf of Y when the distribution of X is $N(0, 1)$.
Ans. $g(y) = \frac{1}{\sqrt{2\pi y}} \exp(-y/2), \quad 0 < y < \infty$.

Example 5.1-5 Let $Y = X^2$. Find the pdf of Y when the distribution of X is

$$f(x) = \frac{x^2}{3}, \qquad -1 < x < 2.$$

Ans. $g(y) = \begin{cases} \frac{\sqrt{y}}{3}, & \text{if } 0 < y < 1 \\ \frac{\sqrt{y}}{6}, & \text{if } 1 < y < 4.. \end{cases}$

Discuss Theorem 5.1-1 about the simulation of r.v.'s.

Exercises from textbook: Section 5.1: 1, 3, 4ab, 5, 10, 11, 15

If X_1 and X_2 are two continuous-type random variables with joint pdf $f(x_1, x_2)$, and if

$$\begin{cases} Y_1 = u_1(X_1, X_2) \\ Y_2 = u_2(X_1, X_2) \end{cases}$$

has the single-valued inverse

$$\begin{cases} X_1 = v_1(Y_1, Y_2), \\ X_2 = v_2(Y_1, Y_2), \end{cases}$$

then the joint pdf of Y_1 and Y_2 is

$$g(y_1, y_2) = |J| \, f\left(v_1(y_1, y_2), v_2(y_1, y_2)\right), \quad (y_1, y_2) \in S_1$$

where the Jacobian J is the determinant

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}.$$

Example 5.2-1 Let X_1 and X_2 be independent random variables, each with pdf

$$f(x) = e^{-x}, \quad 0 < x < \infty.$$

Find the joint pdf of

$$\begin{cases} Y_1 = X_1 - X_2, \\ Y_2 = X_1 + X_2. \end{cases}$$

Find the pdf of Y_1 and Y_2.

Example 5.2-2 Let X and Y be independent uniform r.v.'s over $(0, 1)$. Find the pdf of $Z = XY$.

Recall that if $X_1, X_2, \cdots, X_n$ are independent, then the joint pdf is the product of the respective pdf's (may not be identically distributed), namely,

$$f(x_1, x_2, \cdots, x_n) = f_1(x_1)f_2(x_2)\cdots f_n(x_n).$$

If they follow the same distribution, then:

Definition 5.3-1 *A random sample of size* n refers a collection of *independent and identically distributed (i.i.d.)* random variables $X1, \cdots, X_n$.

Example 5.3-1 Let X_1 and X_2 be independent Poisson random variables with respective means $\lambda_1 = 2$ and $\lambda_2 = 3$.
(a) Find $\mathbb{P}(X_1 = 3, X_2 = 5)$.
(b) Find $\mathbb{P}(X + Y = 1)$.
Ans. (a) 0.0182 (b) 0.0337

Example 5.3-2 An electronic device runs until one of its three components fails. The lifetime (in weeks), X_1, X_2, X_3, of these components are independent, and each has the Weibull pdf

$$f(x) = \frac{2x}{25} e^{-(x/5)^2}, \qquad 0 < x < \infty.$$

Find the probability that the device stops running in the first three weeks.
Ans. 0.660

Theorem 5.3-1 Say $X_1, X_2, \cdots, X_n$ are independent random variables and the random variables $Y = u_1(X_1)u_2(X_2)\cdots u_n(X_n)$. If $\mathbb{E}\left[u_i(X_i)\right], i = 1, 2, \cdots, n$, exists, then

$$\mathbb{E}(Y) = \mathbb{E}\left[u_1(X_1)u_2(X_2)\cdots u_n(X_n)\right] = \mathbb{E}\left[u_1(X_1)\right]\mathbb{E}\left[u_2(X_2)\right]\cdots\mathbb{E}\left[u_n(X_n)\right].$$

Theorem 5.3-2 Let $X_1, X_2, \cdots, X_n$ are independent random variables with respective means $\mu_1, \mu_2 \cdots, \mu_n$ and variances $\sigma_1^2, \sigma_1^2, \cdots, \sigma_n^2$, then the mean and the variance of $Y = \sum_{i=1}^{n} a_i X_i$, where $a_1, a_2, \cdots, a_n$ are real constants, are, respectively,

$$\mu_Y = \sum_{i=1}^{n} a_i \mu_i \quad \text{and} \quad \sigma_Y^2 = \sum_{i=1}^{n} a_i^2 \sigma_i^2 .$$

Corollary 5.3-3 Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from the distribution with mean μ and variance σ^2. Then the *mean of random sample*

$$\overline{X} := \frac{X_1 + X_2 + \cdots + x_n}{n}$$

has the mean and variance as follows:

$$\mu_{\overline{X}} = \sum_{i=1}^{n} \left(\frac{1}{n}\right) \mu = \mu \quad \text{and} \quad \sigma^2_{\overline{X}} = \sum_{i=1}^{n} \left(\frac{1}{n}\right)^2 \sigma^2 = \frac{\sigma^2}{n}.$$

Example 5.3-3 Let X_1 and X_2 be a random sample of size $n = 2$ from the exponential distribution with pdf $f(x) = 2e^{-2x}, 0 < x < \infty$.
(a) Find $\mathbb{P}(0.5 < X_1 < 1.0, 0.7 < X_2 < 1.2)$.
(b) Find $\mathbb{E}[X_1(X_2 - 0.5)^2]$.
Ans. (a) ... (b) ...

Example 5.3-4 Let X_1, X_2, X_3 be three independent random variables with binomial distributions $b(4, 1/2)$, $b(6, 1/3)$, and $b(12, 1/6)$, respectively.
(a) Find $\mathbb{P}(X_1 = 2, X_2 = 2, X_3 = 5)$.
(b) Find $\mathbb{E}(X_1 X_2 X_3)$.
(c) Find the mean and the variance of $Y = X_1 + X_2 + X_3$.
Ans. (a) ... (b) ... (c) ...

Example 5.3-5 Let X_1, X_2, X_3 be independent random variables that represent lifetimes (in hours) of three key components of a device. Say their respective distributions are exponential with means $1000, 1500$, and 2000. Let Y be the minimum of X_1, X_2, X_3 and compute $\mathbb{P}(Y > 1000)$.

Exercises from textbook: Section 5.3: 2, 3, 4, 6, 10, 17, 19.

Theorem 5.4-1 If $X_1, X_2, \cdots X_n$ are independent random variables with respective moment generating functions $M_{X_i}(t), i = 1, 2, 3, \cdots, n$, then the moment-generating function of $Y = \sum_{i=1}^{n} a_i X_i$ is

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(a_i t).$$

In particular, the moment-generating function of $\overline{X} = \sum_{i=1}^{n}(1/n)X_i$ is

$$M_{\overline{X}}(t) = \prod_{i=1}^{n} M_{X_i}(t/n).$$

Example 5.4-1 Let X_1 and X_2 have independent distributions $b(n_1, p)$ and $b(n_2, p)$. Find the moment-generating function of $Y = X_1 + X_2$. How is Y distributed?

Example 5.4-2 Let X_1, X_2, X_3, X_4, X_5 be a random sample of size 5 from a geometric distribution with $p = 1/3$.
(a) Find the moment generating function of $Y = X_1 + X_2 + X_3 + X_4 + X_5$.
(b) How is Y distributed?
(c) Find mgf of $\overline{X}$.
Ans: (a) ... (b) ... (c) ...

Theorem 5.4-2 Let $X_1, X_2, \cdots, X_n$ be independent chi-square random variables with $r_1, r_2, \cdots, r_n$ degrees of freedom, respectively. Then $Y = X_1 + X_2 + \cdots + X_n$ follows $\chi^2(r_1 + r_2 + \cdots + r_n)$.

Corollary 5.4-3 Let $Z_1, Z_2, \cdots, Z_n$ be independent standard normal distributions, $N(0,1)$, random variables, then $W = Z_1^2 + Z_2^2 + \cdots + Z_n^2 \sim \chi^2(n)$.

Corollary 5.4-4 If $X_1, X_2, \cdots, X_n$ be independent and have a normal distributions $N(\mu_i, \sigma_i^2), i = 1, 2, \cdots, n$, respectively, then

$$W = \sum_{i=1}^{n} \frac{(X_i - \mu_i)^2}{\sigma_i^2} \sim \chi^2(n).$$

Exercises from textbook: 5.4.1, 5.4-3, 5.4.4, 5.4-7, 5.4-8, 5.4-15.

Theorem 5.5-1 If $X_1, X_2, \cdots, X_n$ are n mutually independent normal variables with means $\mu_1, \mu_2, \cdots, \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2$, respectively, then the linear function

$$Y = \sum_{i=1}^{n} c_i X_i \sim N\left(\sum_{i=1}^{n} c_i \mu_i, \sum_{i=1}^{n} c_i^2 \sigma_i^2\right).$$

Corollary 5.5-2 If $X_1, X_2, \cdots, X_n$ are observations of a random sample of size n from the normal distribution $N(\mu, \sigma^2)$, then the distribution of the sample mean $\overline{X} = (1/n)\sum_{i=1}^{n} X_i$ follows $N(\mu, \sigma^2/n)$.

Theorem 5.5-3 Let $X_1, X_2, \cdots, X_n$ are observations of a random sample of size n from the normal distribution $N(\mu, \sigma^2)$. Then the *sample mean*

$$\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$$

and the *sample variance*

$$S^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2,$$

are independent. Moreover,

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\sigma^2} \sim \chi^2(n-1).$$

Example 5.5-1 Let X equal the widest diameter (in millimeters) of the fetal head measured between the 16th and 25th weeks of pregnancy. Assume that the distribution of X is $N(\mu = 46.58, \sigma^2 = 40.96)$. Let $\overline{X}$ be the sample mean of a random sample of $n = 16$ observations of X.
(a) Give the value of $\mathbb{E}(\overline{X})$ and $\mathrm{Var}(\overline{X})$.
(b) Find $\mathbb{P}(44.42 \le \overline{X} \le 48.98)$.
(c) How is $W = \sum_{i=1}^{16} \frac{(X_i - \overline{X})^2}{40.96}$ distributed?
(d) Find $\mathbb{P}[6.262 < W < 30.58]$.
Ans: (a) ... (b) ... (c) ... (d) ...

Theorem 5.5-4 (Student's t distribution) Let

$$T = \frac{Z}{\sqrt{U/r}}$$

where Z is a random variable that is $N(0,1)$, U is a random variable that is $\chi^2(r)$, and Z and U are independent. Then the pdf of T is

$$f(t) = \frac{\Gamma((r+1)/2)}{\sqrt{\pi r}\,\Gamma(r/2)} \frac{1}{(1+t^2/r)^{(r+1)/2}}, \quad -\infty < t < \infty.$$

This distribution is called *Student's t distribution*.

We can use the results of Corollary 5.5-2 and Theorem 5.5-3 and Theorem 5.5-4 to construct an important T random variable. Given a random sample $X_1, X_2, \cdots, X_n$ from a normal distribution, $N(\mu, \sigma^2)$, let

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \quad \text{and} \quad U = \frac{(n-1)S^2}{\sigma^2}.$$

Then the distribution of Z is $N(0, 1)$ by Corollary 5.5-2. Theorem 5.5-3 tells us that the distribution of U is $\chi^2(n-1)$ and that Z and U are independent. Thus,

$$T = \frac{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}} = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

has a Student's t distribution with $r = n - 1$ degrees of freedom by Theorem 5.5-4. We use this T to construct confidence intervals for an unknown mean μ of a normal distribution.

Exercises from textbook 5.5-1, 5.5-2, 5.5-3, 5.5-4, 5.5-5, 5.5-6, 5.5-9.

Chapter 5. Distributions of Functions of Random Variables

Theorem 5.6-1 (Central Limit Theorem) If $\overline{X}$ is the mean of a random sample $X_1, X_2, \cdots, X_n$ of size n from a distribution with a finite mean μ and a finite positive variance σ^2, then the distribution of

$$W = \frac{\overline{X} - \mu}{\sqrt{\sigma}/n} = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma}$$

is $N(0, 1)$ in the limit as $n \to \infty$.

When n is "sufficiently large," a practical use of the central limit theorem is approximating the cdf of W, namely,

$$\mathbb{P}(W \leq w) \approx \int_{-\infty}^{w} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = \Phi(w).$$

Example 5.6-1 Let $\overline{X}$ be the mean of a random sample of size 12 from the uniform distribution on the interval $(0, 1)$. Approximate $\mathbb{P}(1/2 \leq \overline{X} \leq 2/3)$.

Example 5.6-2 Let X equal the weight in grams of a miniature candy bar. Assume that $\mu = \mathbb{E}(X) = 24.43$ and $\sigma^2 = Var(X) = 2.20$. Let $\overline{X}$ be the sample mean of a random of $n = 30$ candy bars.

(a) Find $\mathbb{E}\left(\overline{X}\right)$;

(b) Find $\mathrm{Var}\left(\overline{X}\right)$;

(c) Find $\mathbb{P}\left(24.17 \leq \overline{X} \leq 24.82\right)$ approximately.

Ans: (a) ... (b) ... (c) ...

Exercises from textbook: 5.6-2, 5.6-4, 5.6-6, 5.6-7, 5.6-9.

Chapter 5. Distributions of Functions of Random Variables

For sufficiently large n the binomial distribution, $b(n, p)$ can be approximated by normal distribution $N(np, np(1-p))$.

The rule for "sufficiently large" is

$$np \geq 5 \quad \text{and} \quad n(1-p) \geq 5.$$

Example 5.7-1 Let Y be $b(36, 1/2)$. Find $\mathbb{P}(12 < Y \leq 18)$, approximately.
Ans. ≈ 0.5329 and the exact answer is 0.5334.

Chapter 5. Distributions of Functions of Random Variables

Chapter 5. Distributions of Functions of Random Variables