Introduction to Probability and Statistics Welcome Tutorial :-)

Tutorial 10

Lecturer: Ming Gao

DaSE @ ECNU

Nov. 9, 2023

1. a. The cumulative distribution function (CDF) $F_Y(y)$ for a continuous random variable is defined as the integral of its probability density function (pdf) from negative infinity to y:

$$F_Y(y) = \int_{-\infty}^{y} f_Y(t)\,dt$$

Given $f_Y(y) = \frac{1}{\pi(1+y^2)}$, we integrate it to find the CDF:

$$F_Y(y) = \int_{-\infty}^{y} \frac{1}{\pi(1+t^2)}\,dt$$

This integral is a standard form that evaluates to $\frac{\tan^{-1}(t)}{\pi}$. Thus:

$$F_Y(y) = \left[\frac{\tan^{-1}(t)}{\pi}\right]_{-\infty}^{y}$$

We evaluate the integral at the limits:

$$F_Y(y) = \frac{\tan^{-1}(y)}{\pi} - \frac{\tan^{-1}(-\infty)}{\pi}$$

Since $\tan^{-1}(-\infty) = -\frac{\pi}{2}$, we get:

$$F_Y(y) = \frac{\tan^{-1}(y)}{\pi} - \left(-\frac{1}{2}\right) = \frac{1}{2} + \frac{\tan^{-1}(y)}{\pi}$$

b. To simulate a Cauchy random variable from a uniform(0,1) random variable, we use the inverse transform method. The idea is to set the CDF of the desired distribution equal to a uniform(0,1) variable and solve for the variable of interest. For a standard Cauchy distribution $Cauchy(0,1)$, the CDF is $F_Y(y) = \frac{1}{2} + \frac{\tan^{-1}(y)}{\pi}$.
If U is uniform(0,1), we set:

$$U = \frac{1}{2} + \frac{\tan^{-1}(y)}{\pi}$$

Rearranging the equation:

$$\tan^{-1}(y) = \pi\left(U - \frac{1}{2}\right)$$

Taking the tangent of both sides:

$$y = \tan\left(\pi\left(U - \frac{1}{2}\right)\right)$$

This gives a method to simulate a standard Cauchy random variable from a uniform(0,1) random variable. For a general Cauchy distribution with parameters a (location) and b (scale), the transformation is:

$$Y = a + b \cdot \tan\left(\pi\left(U - \frac{1}{2}\right)\right)$$

Here, U is a uniform(0,1) random variable. This transformation will simulate a $Cauchy(a, b)$ random variable.

2. According to the Central Limit Theorem, the sampling distribution of the sample mean $\overline{X}$ will be normally distributed since the population is normally distributed. The mean of the sampling distribution ($\mu_{\overline{X}}$) is equal to the population mean (μ), and the standard deviation of the sampling distribution ($\sigma_{\overline{X}}$), also known as the standard error, is $\sigma/\sqrt{n}$.

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{16}} = \frac{5}{4} = 1.25$$

We convert the interval limits to Z-scores, which are standard normal values:

Lower limit: $\mu_{\overline{X}} - 1.9\sigma_{\overline{X}} = 50 - 1.9 \times 1.25$

Upper limit: $\mu_{\overline{X}} - 0.4\sigma_{\overline{X}} = 50 - 0.4 \times 1.25$

We can then get the Z-scores for the interval limits are as follows:

Lower limit Z-score: -1.9

Upper limit Z-score: -0.4

Now, we will calculate the probability that $\overline{X}$ falls within this interval. This is done by finding the area under the standard normal curve between these two Z-scores.The probability that the sample mean $\overline{X}$ will fall in the interval from $\mu_{\overline{X}} - 1.9\sigma_{\overline{X}}$ to $\mu_{\overline{X}} - 0.4\sigma_{\overline{X}}$ is approximately 0.316.

3. Using the CLT, $\bar{X}$ is approximately $\mathrm{n}\left(\mu, \sigma_{\bar{X}}^2\right)$ with $\sigma_{\bar{X}} = \sqrt{.09} = .3$ and $(\bar{X}-\mu)/.3 \sim \mathrm{n}(0,1)$. Thus

$$.9 = P\left(-1.645 < \frac{\bar{X}-\mu}{.3} < 1.645\right) = P(-.4935 < \bar{X}-\mu < .4935).$$

4. Define a new random variable $X = -\log U$. We need to find the pdf of X.
   The CDF of X, denoted $F_X(x)$, is given by:
   $$F_X(x) = P(X \leq x) = P(-\log U \leq x)$$
   Since U is uniform on $(0,1)$, this can be transformed into:
   $$F_X(x) = P(U \geq e^{-x})$$
   For U uniform on $(0,1)$, this is:
   $$F_X(x) = 1 - e^{-x}, \quad \text{for } x \geq 0$$
   The pdf of X, $f_X(x)$, is the derivative of $F_X(x)$:
   $$f_X(x) = \frac{d}{dx} F_X(x) = e^{-x}, \quad \text{for } x \geq 0$$
   This is the pdf of an exponential random variable with rate 1.

Define a new random variable $Y = -\log(1-U)$. We need to find the pdf of Y. The CDF of Y, denoted $F_Y(y)$, is given by:

$$F_Y(y) = P(Y \leq y) = P(-\log(1-U) \leq y)$$

This can be transformed into:

$$F_Y(y) = P(1-U \geq e^{-y})$$

For U uniform on $(0,1)$, this is:

$$F_Y(y) = P(U \leq 1-e^{-y}) = 1-e^{-y}, \quad \text{for } y \geq 0$$

The pdf of Y, $f_Y(y)$, is the derivative of $F_Y(y)$:

$$f_Y(y) = \frac{d}{dy} F_Y(y) = e^{-y}, \quad \text{for } y \geq 0$$

This is also the pdf of an exponential random variable with rate 1.

5. a. Since $U_j \sim \text{uniform}(0,1)$, the random variable $-\log(U_j)$ is an exponential random variable with rate 1 (as shown in a previous explanation).
   The sum of n independent exponential random variables with rate 1 is a gamma random variable with shape n and scale 1.
   The chi-square distribution with $2n$ degrees of freedom, χ^2_{2n}, is a special case of the gamma distribution with shape n and scale 2.
   Therefore, if $X \sim \text{gamma}(n,2)$, then $X \sim \chi^2_{2n}$.
   Since $-2\sum_{j=1}^{n} \log(U_j)$ is a sum of n exponential random variables scaled by -2, it follows a chi-square distribution with $2n$ degrees of freedom.

b. As in part a, $-\log(U_j)$ is an exponential random variable with rate 1. The sum of n such independent exponential random variables is a gamma random variable with shape n and scale 1. If we multiply an exponential random variable with rate 1 by β, it becomes an exponential random variable with rate $1/\beta$. Therefore, $-\beta\sum_{j=1}^{n}\log(U_j)$ is a gamma random variable with shape n and scale β.

c. As shown earlier, $-\sum_{j=1}^{n} \log(U_j)$ and $-\sum_{j=1}^{n+m} \log(U_j)$ are gamma distributed with shapes n and $n+m$, and both with scale 1.
The ratio of two independent gamma random variables, where the numerator has a shape parameter n and the denominator has a shape parameter $n+m$ (both with the same scale), follows a beta distribution with parameters n and m.
Therefore, $\frac{\sum_{j=1}^{n} \log(U_j)}{\sum_{j=1}^{n+m} \log(U_j)}$ follows a beta distribution with parameters n and m.