Statistics Preliminaries

Lorenzo Naranjo

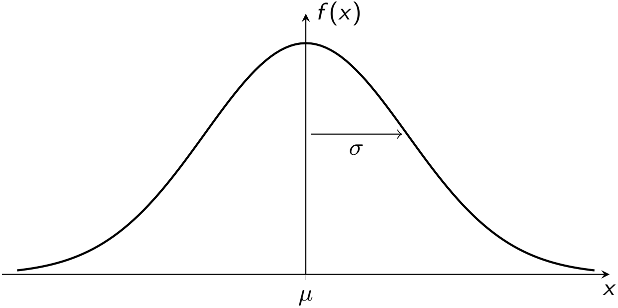
WashU Olin
Business School

The Normal Distribution

- We say that a random variable (RV) X is normally distributed with mean μ and standard deviation σ if its probability density function is:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

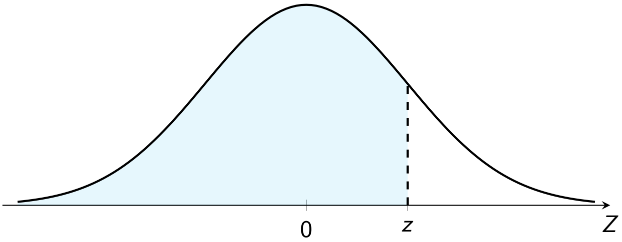
and we usually write $X \sim \mathrm{N}(\mu, \sigma^2)$.



Standard Normal Cumulative Probability

- If $Z \sim \mathrm{N}(0,1)$, we have that:

$$\Pr(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx = \Phi(z)$$



- Since the integral cannot be solved in closed-form, the probability must then be obtained from a table or using a computer.

Computing Normal Probabilities

- In order to compute cumulative probabilities for $X \sim \mathrm{N}(\mu, \sigma^2)$, we usually normalize the random variable X:

$$\Pr(X \leq x) = \Pr\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = \Pr\left(Z \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

where $Z = \frac{X-\mu}{\sigma} \sim \mathrm{N}(0,1)$ is called a Z-score.

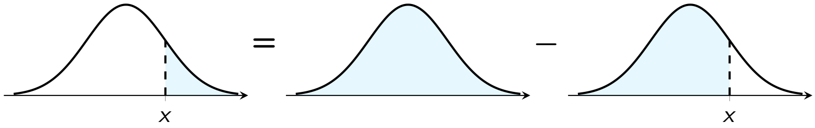
Example 1

Suppose that $X \sim \mathrm{N}(\mu, \sigma^2)$ with $\mu = 10$ and $\sigma = 25$. What is the probability that $X \leq 0$?

$$\Pr(X \leq 0) = \Pr\left(Z \leq \frac{0-10}{25}\right) = \Phi(-0.40) = 0.3446.$$

Computing a Right-Tail Probability

- Right-Tail Probability: $\Pr(X > x) = 1 - \Pr(X \leq x)$



Example 2

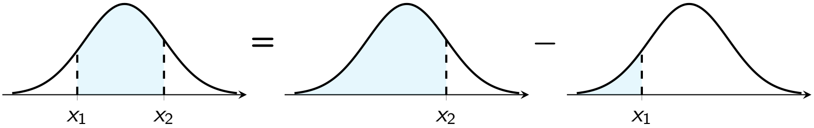
Suppose that $X \sim \mathrm{N}(\mu, \sigma^2)$ with $\mu = 10$ and $\sigma = 25$. What is the probability that $X > 12$?

$$\Pr(X \leq 12) = \Pr\left(Z \leq \tfrac{12-10}{25}\right) = \Phi(0.08) = 0.5319.$$

Therefore, $\Pr(X > 12) = 1 - 0.5319 = 0.4681$.

Computing an Interval Probability

- Interval Probability: $\Pr(x_1 < X \leq x_2) = \Pr(X \leq x_2) - \Pr(X \leq x_1)$



Example 3

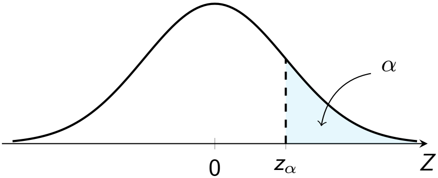
Suppose that $X \sim N(\mu, \sigma^2)$ with $\mu = 10$ and $\sigma = 25$. What is the probability that $2 < X \leq 14$?

$$\Pr(X \leq 14) = \Pr\left(Z \leq \tfrac{14-10}{25}\right) = \Phi(0.16) = 0.5636,$$
$$\Pr(X \leq 2) = \Pr\left(Z \leq \tfrac{2-10}{25}\right) = \Phi(-0.32) = 0.3745.$$

Therefore, $\Pr(2 < X \leq 14) = 0.5636 - 0.3745 = 0.1891$.

Computing a Standard Normal Right-Tail Percentile

- For a standard normal variable Z, a right-tail percentile is the value z_α above which we obtain a certain probability α.



- Mathematically, this means finding z_α such that:

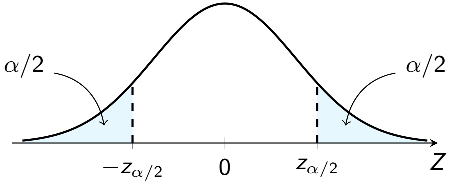
$$\Pr(Z > z_\alpha) = \alpha \Leftrightarrow \Pr(Z \leq z_\alpha) = 1 - \alpha$$

- The following table shows common values for z_α:

α	0.05	0.025	0.01	0.005
z_α	1.64	1.96	2.33	2.58

Confidence Interval for a Standard Normal RV

- A $1 - \alpha$ confidence interval (CI) defines left and right percentiles such that the probability on each side is $\alpha/2$.



- For a standard normal variable Z, the symmetry of its PDF implies:

$$\Pr(Z \leq -z_{\alpha/2}) = \Pr(Z > z_{\alpha/2}) = \alpha/2$$

Example 4

Since $z_{2.5\%} = 1.96$, the 95% confidence interval of Z is $[-1.96, 1.96]$. This means that if we randomly sample this variable 100,000 times, approximately 95,000 observations will fall inside this interval.

Confidence Interval for a Normal RV

- If $X \sim \mathrm{N}(\mu, \sigma^2)$, its confidence interval is determined by ξ and ζ such that:

$$\Pr(X \leq \xi) = \alpha/2 \Rightarrow \Pr(Z \leq \tfrac{\xi-\mu}{\sigma}) = \alpha/2,$$
$$\Pr(X > \zeta) = \alpha/2 \Rightarrow \Pr(Z > \tfrac{\zeta-\mu}{\sigma}) = \alpha/2,$$

  which implies that $-z_{\alpha/2} = \frac{\xi-\mu}{\sigma}$ and $z_{\alpha/2} = \frac{\zeta-\mu}{\sigma}$.
- The $1-\alpha$ confidence interval for X is then $[\mu - z_{\alpha/2}\sigma, \mu + z_{\alpha/2}\sigma]$.

Example 5

Suppose that $X \sim \mathrm{N}(\mu, \sigma^2)$ with $\mu = 10$ and $\sigma = 25$. Since $z_{2.5\%} = 1.96$, the 95% confidence interval of X is:

$$[10 - 1.96(25), 10 + 1.96(25)] = [-39, 59].$$

The Lognormal Distribution

- If $X \sim N(\mu, \sigma^2)$ then $Y = e^X$ is said to be log-normally distributed with the same parameters.
- The distribution function of the lognormal distribution can be obtained as follows:

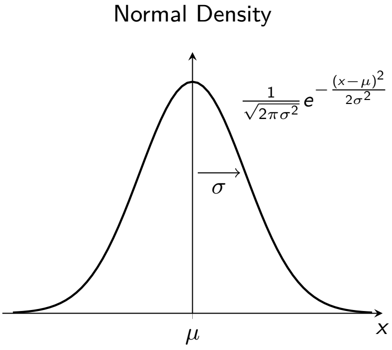
$$\begin{aligned}\Pr(Y \leq y) &= \Pr(X \leq \ln(y)) \\ &= \int_{-\infty}^{\ln(y)} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx\end{aligned}$$

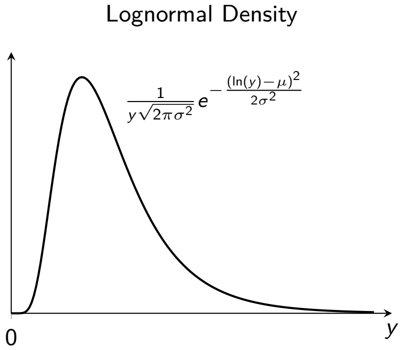
- We can then define $z = e^x \Leftrightarrow x = \ln(z) \Rightarrow dx = (1/z)dz$ and

$$\Pr(Y \leq y) = \int_{-\infty}^{y} \frac{1}{z\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln(z)-\mu)^2}{2\sigma^2}}\, dz$$

Probability Density Function Comparison

- Unlike the normal density, the lognormal density function is not symmetric around its mean.
- Normally distributed variables can take values in $(-\infty, \infty)$, whereas lognormally distributed variables are always positive.





Computing Probabilities for Lognormal RVs

- We can use the fact that the logarithm of a lognormal random variable is normally distributed to compute cumulative probabilities.

Example 6

Let $Y = e^{4+1.5Z}$ where $Z \sim \mathrm{N}(0,1)$. What is the probability that $Y \leq 100$?

$$\begin{aligned} \Pr(Y \leq 100) &= \Pr(e^X \leq 100) \\ &= \Pr(X \leq \ln(100)) \\ &= \Pr\left(Z \leq \tfrac{\ln(100)-4}{1.5}\right) \\ &= \Phi(0.4034) \\ &= 0.6567 \end{aligned}$$

Confidence Interval for a Lognormal RV

- Let $Y = e^{\mu+\sigma Z}$ where $Z \sim \mathrm{N}(0,1)$. We have that:

$$
\begin{aligned}
-z_{\alpha/2} < Z \leq z_{\alpha/2} &\Rightarrow \mu - \sigma z_{\alpha/2} < \mu + \sigma Z \leq \mu + \sigma z_{\alpha/2} \\
&\Rightarrow e^{\mu - \sigma z_{\alpha/2}} < e^{\mu+\sigma Z} \leq e^{\mu + \sigma z_{\alpha/2}}
\end{aligned}
$$

- The $1-\alpha$ confidence interval for Y is $[e^{\mu-\sigma z_{\alpha/2}}, e^{\mu+\sigma z_{\alpha/2}}]$.

Example 7

Let $Y = e^{4+1.5Z}$ where $Z \sim \mathrm{N}(0,1)$. The 95% confidence interval for Y is:

$$[e^{4-1.96(1.5)}, e^{4+1.96(1.5)}] = [2.89, 1032.71].$$

Moments of a Lognormal Random Variable

- Let $Y = e^{\mu+\sigma Z}$ where $Z \sim \mathrm{N}(0,1)$. We have that:

$$\mathrm{E}(Y) = e^{\mu+0.5\sigma^2} \tag{1}$$

$$\mathrm{V}(Y) = e^{2\mu+\sigma^2}(e^{\sigma^2}-1) \tag{2}$$

$$\mathrm{SD}(Y) = \mathrm{E}(Y)\sqrt{e^{\sigma^2}-1} \tag{3}$$

Example 8

Let $Y = e^{4+1.5Z}$ where $Z \sim \mathrm{N}(0,1)$. The expectation and standard deviation of Y are:

$$\mathrm{E}(Y) = e^{4+0.5(1.5^2)} = 168.17$$

$$\mathrm{SD}(Y) = 168.17\sqrt{e^{1.5^2}-1} = 489.95$$

Partial Expectations

- When pricing a call option, the payoff is positive if the option is in-the-money and zero otherwise.
- We usually use an indicator function to quantify this behavior:

$$\mathbb{1}_{\{Y>K\}} = \begin{cases} 0 & \text{if } Y \leq K \\ 1 & \text{if } Y > K \end{cases}$$

- Let $Y = e^X$ where $X \sim \mathrm{N}(\mu, \sigma^2)$. Then we have that:

$$\mathrm{E}\left(Y \mathbb{1}_{\{Y>K\}}\right) = e^{\mu + \frac{1}{2}\sigma^2}\,\Phi\left(\frac{\mu + \sigma^2 - \ln(K)}{\sigma}\right) \quad (4)$$

$$\mathrm{E}\left(K \mathbb{1}_{\{Y>K\}}\right) = K\,\Phi\left(\frac{\mu - \ln(K)}{\sigma}\right) \quad (5)$$