Math 3070, Applied Statistics

Section 1

September 30, 2019

Lecture Outline, 9/30

Section 4.5 and 4.6

- Gamma Function
- Weibull Distribution
- Lognormal Distribution
- Beta Distribution
- Probability Plots

Gamma Function, Summary

- $\Gamma(k) = \int_0^\infty x^{k-1} e^{-x} dx$
- $\Gamma(1) = 1$
- $\Gamma(1/2) = \sqrt{\pi}$
- $\Gamma(k+1) = k\Gamma(k), \quad k > 0$
- $\Gamma(k+1) = k!, \quad \text{integer } k$

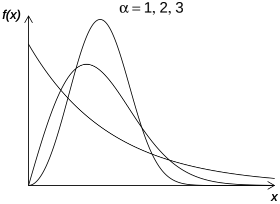
$$\begin{aligned}
\Gamma(k+1) &= \int_0^\infty x^k e^{-x} dx \\
&= -x^k e^{-x} \Big|_{x=0}^{\infty} + k \int_0^\infty x^{k-1} e^{-x} dx \\
&= 0 + k\Gamma(k)
\end{aligned}$$

Weibull Distribution, PDF

A **Weibull** random variable X with shape $\alpha > 0$ and scale $\beta > 0$ has pdf

$$f(x) = \begin{cases} \frac{\alpha}{\beta^\alpha} x^{\alpha - 1} e^{-(x/\beta)^\alpha}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

$X \sim \mathit{Weibull}(\alpha, \beta)$



Important for central limit theorems involving extreme values. For example, $\max(X_1, X_2, \ldots X_n)$ or $\min(X_1, X_2, \ldots X_n)$ as $n \to \infty$.

Weibull Distribution, Check PDF

Suppose $X \sim \textit{Weibull}(\alpha, \beta)$.
$f(x) \geq 0$.

$$\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{\infty} \frac{\alpha}{\beta^{\alpha}} x^{\alpha-1} e^{-(x/\beta)^{\alpha}} dx$$

$$\text{use } u = x^{\alpha} \to du = \alpha x^{\alpha-1} dx$$

$$= \int_{0}^{\infty} \frac{e^{-x^{\alpha}/\beta^{\alpha}}}{\beta^{\alpha}} \alpha x^{\alpha-1} dx$$

$$= \int_{0}^{\infty} \frac{e^{-u/\beta^{\alpha}}}{\beta^{\alpha}} du$$

$$= -e^{-u/\beta^{\alpha}} \Big|_{u=0}^{\infty} = -e^{-(x/\beta)^{\alpha}} \Big|_{x=0}^{\infty}$$

$$= 0 - (-1) = 1$$

$f(x)$ is a PDF.

Weibull Distribution, CDF

Suppose $X \sim \mathit{Weibull}(\alpha, \beta)$. The previous calculation shows us how to calculate the CDF.

$$
\begin{aligned}
F(x) = P(X < x) = \int_{-\infty}^{x} f(y)dy &= \int_{0}^{x} \frac{\alpha}{\beta^\alpha} y^{\alpha-1} e^{-(y/\beta)^\alpha} dy \\
&= -e^{-(y/\beta)^\alpha} \Big|_{y=0}^{x} \\
&= -e^{-(x/\beta)^\alpha} - (-1) \\
&= 1 - e^{-(x/\beta)^\alpha}
\end{aligned}
$$

Weibull Distribution, Expected Value

Suppose $X \sim \mathit{Weibull}(\alpha, \beta)$.

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx = \int_0^{\infty} x\frac{\alpha}{\beta^\alpha}x^{\alpha-1}e^{-(x/\beta)^\alpha}dx$$

$$\text{use } u = (x/\beta)^\alpha \to du = \alpha\frac{x^{\alpha-1}}{\beta^\alpha}dx \text{ and } x = \beta u^{1/\alpha}$$

$$= \int_0^{\infty} xe^{-u}du = \beta\int_0^{\infty} u^{1/\alpha}e^{-u}du$$

$$= \beta\int_0^{\infty} u^{(1+1/\alpha)-1}e^{-u}du$$

$$= \beta\Gamma\left(1+\frac{1}{\alpha}\right)$$

Weibull Distribution, Variance

Suppose $X \sim \mathit{Weibull}(\alpha, \beta)$.
Use $\mathit{Var}(X) = E[X^2] - E[X]^2$

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_0^{\infty} x^2 \frac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-(x/\beta)^\alpha} dx$$

$$\text{use } u = (x/\beta)^\alpha \rightarrow du = \alpha \frac{x^{\alpha-1}}{\beta^\alpha} dx \text{ and } x = \beta u^{1/\alpha}$$

$$= \int_0^{\infty} x^2 e^{-u} du = \beta^2 \int_0^{\infty} u^{2/\alpha} e^{-u} du$$

$$= \beta^2 \int_0^{\infty} u^{(1+2/\alpha)-1} e^{-u} du$$

$$= \beta^2 \Gamma\left(1 + \frac{2}{\alpha}\right)$$

$$V(X) = E[X^2] - E[X]^2 = \beta^2 \left[\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma\left(1 + \frac{1}{\alpha}\right)^2\right]$$

Weibull Distribution, Expected Value Example

Suppose that $X \sim \mathit{Weibull}(2, \beta)$ and $E[X] = 3$. Determine β, the scale parameter.

$$
\begin{aligned}
3 = E[X] &= \beta \Gamma\left(1 + \frac{1}{2}\right) \\
&= \beta \frac{1}{2} \Gamma\left(\frac{1}{2}\right) \\
&= \beta \frac{1}{2} \sqrt{\pi} \\
\rightarrow \beta &= \frac{6}{\sqrt{\pi}}
\end{aligned}
$$

Weibull Distribution, Typical Example

Suppose that $X \sim \mathit{Weibull}(2, 3)$. What is the probability that X is greater than 3.

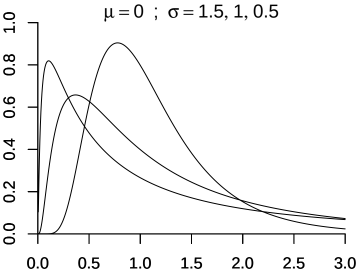
$$\begin{aligned} P(X > 3) &= 1 - P(X \leq 3) \\ &\text{use the CDF} \\ &= 1 - \left(1 - e^{-(3/\beta)^{\alpha}}\right) \\ &= e^{-(3/3)^2} = e^{-1} \end{aligned}$$

Weibull Distribution, Summary

- $X \sim \mathit{Weibull}(\alpha, \beta)$
- $$f(x) = \begin{cases} \frac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-(x/\beta)^\alpha}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$
- $$F(x) = \begin{cases} 1 - e^{-(x/\beta)^\alpha}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$
- $$E[X] = \beta\Gamma\left(1 + \frac{1}{\alpha}\right) \qquad V(X) = \beta^2\left[\Gamma\left(1 + \frac{2}{\alpha}\right) - \left(1 + \frac{1}{\alpha}\right)^2\right]$$
- Think of this if you need a central limit theorem for extreme values later in life.
`https://en.wikipedia.org/wiki/Extreme_value_theory`

A random variable X is said to have a **log-normal** distribution if $\ln(X)$ is a normal random variable $N(\mu, \sigma)$.
$X \sim Lognormal(\mu, \sigma)$

A log-normal random variable X may be written in the form $X = e^{\sigma Z + \mu}$, where Z is a standard normal random variable.



Used with Black-Scholes and other compound rate models.
Note, $\propto$ should be μ in the plot.

Log-normal Distribution, CDF

Suppose $X \sim Lognormal(\mu, \sigma)$.

$$
\begin{aligned}
F(x) = P(X < x) &= P(e^{\sigma Z + \mu} < x) \\
&= P\left(Z < \frac{\ln(x) - \mu}{\sigma}\right) \\
&= \Phi\left(\frac{\ln(x) - \mu}{\sigma}\right)
\end{aligned}
$$

range of log all real numbers and domain is $(0, \infty)$

$(0, \infty]$ are the possible values of X

$$
F(x) = \begin{cases} \Phi\left(\frac{\ln(x)-\mu}{\sigma}\right), & x > 0 \\ 0, & x \leq 0 \end{cases}
$$

Log-normal Distribution, PDF

Suppose $X \sim \textit{Lognormal}(\mu, \sigma)$.
Differentiate for PDF.
$f(x) = 0$ when $x \leq 0 \rightarrow F(x) = 0$ when $x \leq 0$ When $x > 0$,

$$\begin{aligned}
\frac{d}{dx}\Phi\left(\frac{\ln(x) - \mu}{\sigma}\right) &= \Phi'\left(\frac{\ln(x) - \mu}{\sigma}\right)\left[\frac{d}{dx}\frac{\ln(x) - \mu}{\sigma}\right] \\
&= \frac{1}{\sigma x}\Phi'\left(\frac{\ln(x) - \mu}{\sigma}\right) \\
&= \frac{1}{\sigma x}\frac{1}{\sqrt{2\pi}}\exp\left(\frac{-[\ln(x) - \mu]^2}{2\sigma^2}\right)
\end{aligned}$$

Is a PDF since the CDF was known.

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma x}\exp\left(\frac{-[\ln(x) - \mu]^2}{2\sigma^2}\right), & x > 0 \\ 0, & x \leq 0 \end{cases}$$

Log-normal Distribution, Mean

Suppose $X \sim Lognormal(\mu, \sigma)$.

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx = \int_{0}^{\infty} \frac{x}{\sqrt{2\pi}} \exp\left(\frac{-[\ln(x)-\mu]^2}{2\sigma^2}\right)\frac{dx}{\sigma x}$$

use u-sub: $y = \dfrac{\ln(x)-\mu}{\sigma} \rightarrow x = e^{\sigma y+\mu},\ dy = \dfrac{dx}{\sigma x}$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(\sigma y + \mu)\exp(-y^2/2)dy$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\left[\frac{y^2}{2} - \frac{\sigma y}{2} + \frac{\sigma^2}{2}\right] + \mu + \frac{\sigma^2}{2}\right)dy$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-(y-\sigma)^2}{2}\right)\exp\left(\mu + \frac{\sigma^2}{2}\right)dy$$

$$= \exp\left(\mu + \frac{\sigma^2}{2}\right)\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-(y-\sigma)^2}{2}\right)dy$$

$$= \exp(\mu + \sigma^2/2),\ \text{used} \int_{-\infty}^{\infty} f(x)dx = 1 \text{ for } N(\sigma, 1)$$

Log-normal Distribution, Variance

Suppose $X \sim \mathit{Lognormal}(\mu, \sigma)$.
Use the same trick and $V(X) = E[X^2] - E[X]^2$

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_0^{\infty} \frac{x^2}{\sqrt{2\pi}} \exp\left(\frac{-[\ln(x) - \mu]^2}{2\sigma^2}\right)\frac{dx}{\sigma x}$$

use u-sub: $y = \dfrac{\ln(x) - \mu}{\sigma} \to x = e^{\sigma y + \mu},\ dy = \dfrac{dx}{\sigma x}$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(2\sigma y + 2\mu) \exp(-y^2/2) dy$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\left[\frac{y^2}{2} - \frac{4\sigma y}{2} + \frac{4\sigma^2}{2}\right] + 2\mu + 2\sigma^2\right) dy$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-(y - 2\sigma)^2}{2}\right) \exp(2\mu + 2\sigma^2) dy$$

$$= \exp(2\mu + 2\sigma^2),\ \text{used} \int_{-\infty}^{\infty} f(x)dx = 1 \text{ for } N(2\sigma, 1)$$

$$V(X) = e^{2\sigma^2 + 2\mu} - e^{\sigma^2 + 2\mu} = e^{2\mu + \sigma^2}[e^{\sigma^2} - 1]$$

Log-normal Distribution, Typical Example

$X \sim \textit{Lognormal}(1, 2)$.
Compute the probability X is greater than e^2.

$$
\begin{aligned}
P(X \geq e^2) &= 1 - P(X < e^2) \\
P(X < e^2) &= P(\ln(X) < \ln[e^2]) = P(\ln(X) < 2) \\
&= P\left(\frac{\ln(X) - 1}{2} < \frac{2-1}{2}\right) \\
&= P\left(Z < \frac{1}{2}\right), \quad Z \sim N(0,1) \\
&= \Phi\left(\frac{1}{2}\right) \approx 0.6915 \\
P(X \geq e^2) &\approx 1 - 0.6915 = 0.3085
\end{aligned}
$$

Log-normal Distribution, Normal Example (Pun Intended)

Internet browsing time X is modeled by a log normal distribution. One researcher claims that $\ln(X)$ has a mean of 5 log minutes while another claims that X has an expected value of 150 minutes. Can both claims be true?

Check if these parameters are valid.

$$\begin{aligned}
5 = E[\ln(X)] &= \mu \\
150 = E[X] &= e^{\mu+\sigma^2/2} \\
&= e^{5+\sigma^2/2} \\
\ln(150) &= 5 + \sigma^2/2 \\
\sigma^2 &= 2 * (\ln(150) - 5) \\
&\approx -0.11671515478 < 0
\end{aligned}$$

Variances cannot be negative. This situation is impossible. Moreover, the expected value of two related quantities revealed variance of one of them. Ain't that something?

Log-Normal, Summary

- $X \sim Lognormal(\mu, \sigma) \iff \ln(X) \sim N(\mu, \sigma)$
- PDF
$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(\frac{-[\ln(x)-\mu]^2}{2\sigma^2}\right), & x > 0 \\ 0, & x \leq 0 \end{cases}$$
- CDF is not given. Related it to Φ using natural log.
-
$$E[X] = e^{\mu+\sigma^2/2} \qquad V(X) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$$

Beta Distribution

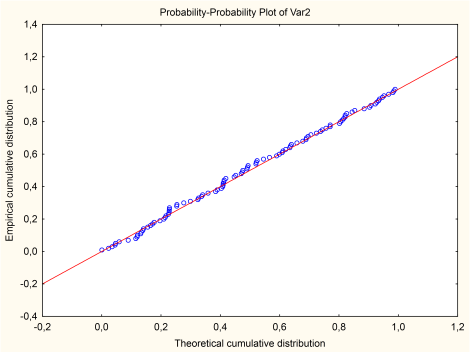
X is a **Beta** random variable with parameters α, β, A and B if it has the following PDF,

$$f(x) = \begin{cases} \dfrac{1}{B-A} \cdot \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)} \left(\dfrac{x-A}{B-A}\right)^{\alpha-1} \left(\dfrac{B-x}{B-A}\right)^{\beta-1}, & A \le x \le B \\ 0, & \text{otherwise} \end{cases}$$

- Nonzero on a bounded interval.
- Sometimes appears when modeling random ratios. For example, the ratio of two gammas.
- Will not ask anything about it in this course. No, not even extra credit. If someone asks, we briefly discussed it. Feel free to ask me about it.

Probability Plots

Goal: assess if data follows a theoritical distribution.
Idea: Plot the percentiles of the data (y-axis) versus the theoritical percentiles (x-axis) to see if they match.



Note: Linear relationships in the plot translate to changes in mean and variance, suggesting ways to adjust parameters. Will not test over probability plots. Will appear in R Lab.