Lecture 3
Common Families of Distributions

- In this chapter, we introduce a variety of discrete and continuous probability distributions that are commonly used in economics and finance.
- Examples of discrete probability distributions include Bernoulli, Binomial, Negative Binomial, Poisson, and Geometric Distributions.
- Examples of continuous probability distributions include uniform, Beta, Exponential, Gamma, normal, lognormal, Cauchy, and generalized Gamma distributions.
- The properties of these distributions as well as their applications in economics and finance are discussed. We also show some important techniques of obtaining moments for various probability distributions.

- The probability space $(S, \mathcal{F}, P)$ completely characterizes a random experiment. In practice, the true probability distribution for the underlying random experiment is usually unknown.
- One usually consider a class of probability measures, say pmf/pdf $f_X(x, \theta)$, each of which is indexed by a parameter value.
- One main objective of statistics and econometrics is to use the observed economic data to estimate the true parameter value, say θ_0, under the assumption that there exists some parameter value θ_0 such that the density model $f_X(x, \theta_0)$ coincides with true probability distribution $g(x)$ of the underlying random experiment.
- Before we come to the stage of estimating the true model parameter (Chapter 7), we first introduce a number of important parametric distribution models and discuss their properties and applications in economics and finance.

1. Discrete Probability Distributions

Bernoulli Distribution

Consider a random experiment where the outcome "success" occurs with probability p $(0 \leq p \leq 1)$ and the outcome "failure" occurs with probability $1 - p$. The probability space of this random experiment is $\{S, \mathcal{F}, P\}$, where

$$\begin{aligned} S &= \{\text{failure, success}\} \\ \mathcal{F} &= \{\emptyset, S, \{\text{failure}\}, \{\text{success}\}\} \\ P(\emptyset) &= 0, P(S) = 1, P(\{\text{failure}\}) = 1 - p, P(\{\text{success}\}) = p, \end{aligned}$$

where $0 < p < 1$. This random experiment is known as a Bernoulli trial.

Define the random variable X by

$$X(\text{failure}) = 0, \;\; X(\text{success}) = 1.$$

Then $\Pr(X = 0) = 1 - p$ and $\Pr(X = 1) = p$.

Theorem: For a Bernoulli(p) random variable, we have $\mathbb{E}(X) = p$ and $Var(X) = p(1-p)$. All higher moments of X are a function of p. $\square$

Exercise: Show this. CB p. 89.

Example: This binary distribution can arise when one tosses a coin whose head shows up with probability p: It has wide applications in economics and finance. For example, one can define a random variable X to take value 1 if the IBM stock price goes up, and to take value 0 if the IBM stock price goes down. Then X is the directional indicator of the IBM stock price and follows a Bernoulli distribution. In many economic applications, the interest is to explain p using a set of economic variables (covariates) Z. For example, one can explain p using

$$\Pr(X = 1|Z) = \frac{1}{1 + \exp(-Z'\beta)}.$$

This is called a logit model, which has been a popular econometric model for binary choice problems where the outcome only has two possibilities.

Binomial Distribution

Next, let us repeat this Bernoulli random experiment n times. For the jth experiment we can define the random variable X_j in the same way as X. Assuming that the random experiments involved are carried out independently from each other, we get a sequence $X_1, \ldots, X_n$ of *independent* random variables each distributed as X. The random variable Y_n defined by

$$Y_n = \sum_{j=1}^{n} X_j$$

is the number of successes in the n trials of the experiment and Y_n has the Binomial(n, p) distribution, with density

$$f_Y(y) = \Pr(Y_n = y) = \left(\begin{array}{c} n \\ y \end{array} \right) p^y (1-p)^{n-y}, \qquad y = 0, 1, 2, \ldots, n.$$

Remark. $f_Y(y)$ is a density as $f_Y(y) \geq 0$ for all y and $\sum_{y=0}^{n} f_Y(y) = 1$:

$$
\begin{aligned}
(a+b)^n &= \sum_{i=0}^{n} \binom{n}{i} b^i a^{n-i} \\
\sum_{y=0}^{n} f_Y(y) &= \sum_{y=0}^{n} \binom{n}{y} p^y (1-p)^{n-y} = [p + (1-p)]^n = 1
\end{aligned}
$$

Example: The sum of independent Bernoulli trials follows the binomial distribution. The binomial distribution has wide applications in economics.

- It can be used to approximate the distribution of the numbers of defective products in a total of n outputs.
- It can also be used to model jumps in financial price movements (e.g., Gonzalez-Rivera, Lee and Mishra 2008 *Journal of Applied Econometrics*).

Theorem: If Y is distributed Binomial(n, p), then $\mathbb{E}Y = np$ and $Var(Y) = np(1-p)$. $\square$

Proof 1: Prove the theorem using pmf. See CB Examples 2.2.3 and 2.3.5.

Proof 2: Prove the theorem using mgf.

$$\begin{aligned} M(t) &= \sum_{y=0}^{n} e^{ty} f_Y(y) = \sum_{y=0}^{n} e^{ty} \begin{pmatrix} n \\ y \end{pmatrix} p^y (1-p)^{n-y} \\ &= \sum_{y=0}^{n} \begin{pmatrix} n \\ y \end{pmatrix} \left(pe^t\right)^y (1-p)^{n-y} = [pe^t + (1-p)]^n \end{aligned}$$

$$\begin{aligned} M^{(1)}(t)|_{t=0} &= n[pe^t + (1-p)]^{n-1} pe^t \;|_{t=0} = np = \mu \\ M^{(2)}(t)|_{t=0} &= (n-1)n[pe^t + (1-p)]^{n-1} \left(pe^t\right)^2 \\ &\quad + n[pe^t + (1-p)]^{n-1} pe^t \;|_{t=0} \\ &= (n-1)np^2 + np \end{aligned}$$

$$
\begin{aligned}
\sigma^2 &= \left[(n-1)np^2 + np\right] - (np)^2 \\
&= n^2p^2 - np^2 + np - n^2p^2 \\
&= np(1-p)
\end{aligned}
$$

Proof 3: Prove the theorem using the cumulant generating function.

Proof 4: Prove the theorem using the characteristic function.

Poisson Distribution

PMF of Poisson(λ) :
A d.r.v. X follows a Poisson(λ) distribution if its pmf

$$f_X(x|\lambda) = e^{-\lambda}\frac{\lambda^x}{x!}, \qquad x = 0, 1, 2, \ldots$$

where $\lambda > 0$. The parameter λ is called an *intensity* parameter.
We first verify that $\sum_{x=0}^{\infty} f_X(x|\lambda) = 1$ for any given $\lambda > 0$.
Recalling the series expansion (Hamilton 1994, p. 716)

$$e^{\lambda} = \sum_{x=0}^{\infty}\frac{\lambda^x}{x!},$$

we have

$$\sum_{x=0}^{\infty} f_X(x|\lambda) = e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} = 1.$$

Mean of Poisson(λ) :

$$\begin{aligned}\mu_X &= \sum_{x=0}^{\infty} x f_X(x|\lambda) \\ &= \sum_{x=0}^{\infty} x e^{-\lambda}\frac{\lambda^x}{x!} \\ &= \sum_{x=1}^{\infty} x e^{-\lambda}\frac{\lambda^x}{x!} \\ &= \lambda \sum_{x=1}^{\infty} e^{-\lambda}\frac{\lambda^{x-1}}{(x-1)!} \\ &= \lambda \sum_{y=0}^{\infty} e^{-\lambda}\frac{\lambda^y}{y!} \qquad\qquad (y = x-1) \\ &= \lambda\end{aligned}$$

where the last sum can be viewed as the sum of the pmf of a random variable $Y \sim$ Poisson(λ).

Variance of Poisson(λ):

$$
\begin{aligned}
\mathbb{E}(X^2) &= \sum_{x=0}^{\infty} x^2 e^{-\lambda} \frac{\lambda^x}{x!} \\
&= \sum_{x=1}^{\infty} x e^{-\lambda} \frac{\lambda^x}{(x-1)!} \\
&= \lambda \sum_{x=1}^{\infty} x e^{-\lambda} \frac{\lambda^{x-1}}{(x-1)!} \\
&= \lambda \sum_{y=0}^{\infty} (y+1) e^{-\lambda} \frac{\lambda^y}{y!} \\
&= \lambda \sum_{y=0}^{\infty} y e^{-\lambda} \frac{\lambda^y}{y!} + \lambda \sum_{y=0}^{\infty} e^{-\lambda} \frac{\lambda^y}{y!} \\
&= \lambda^2 + \lambda.
\end{aligned}
$$

Therefore the variance is

$$\sigma_X^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = (\lambda^2 + \lambda) - \lambda^2 = \lambda.$$

So, the parameter λ is both the mean and the variance of the Poisson distribution.

MGF of Poisson(λ) :

$$
\begin{aligned}
M_X(t) &= \mathbb{E}(e^{tX}) \\
&= \sum_{x=0}^{\infty} e^{tx} f_X(x|\lambda) \\
&= \sum_{x=0}^{\infty} e^{tx} e^{-\lambda} \frac{\lambda^x}{x!} \\
&= e^{-\lambda} \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x}{x!} \\
&= e^{-\lambda} \sum_{x=0}^{\infty} \frac{\left(\lambda e^t\right)^x}{x!} \\
&= e^{-\lambda} \cdot e^{\left(\lambda e^t\right)} \qquad \left[\text{using } e^a = \sum_{x=0}^{\infty} \frac{a^x}{x!}\right] \\
&= e^{\lambda(e^t - 1)}.
\end{aligned}
$$

See CB Exercise 2.33 and Example 2.3.13.

Remark: Why is the Poisson distribution useful in practice?

- It has been used to model the number of events (e.g., number of customers passing through a cashier counter, number of telephone calls, number of jumps in a financial price) in a unit of time.
- One of the basic assumptions on which the Poisson distribution is built is that, for small time intervals, the probability of an arrival is proportional to the length of waiting time. This makes it a reasonable model for situations like those mentioned above.

Example (Waiting time) CB Example 3.2.4 . Consider a telephone operator who, on average, handles five calls every 3 minutes. What is the probability that there will be no calls in the next minute? At least two calls? Note that $\lambda = 5/3$. Find $\Pr(X = 0)$ and $\Pr(X \geq 2)$.

Poisson approximation to the binomial distribution

Recall the mgf of the binomial distribution $B(n,p)$ is

$$M_B(t) = \left[pe^t + (1-p)\right]^n,$$

and the mgf of the Poisson distribution $P(\lambda)$ is

$$M_P(t) = e^{\lambda(e^t-1)}.$$

When $n \to \infty$ but $np \to \lambda$, we can use the Poisson distribution to approximate the binomial distribution because

$$\begin{aligned} M_B(t) &= \left[pe^t + (1-p)\right]^n \\ &= \left[1 + \frac{np(e^t-1)}{n}\right]^n \\ &\to e^{\lambda(e^t-1)} = M_P(t), \end{aligned}$$

where the last line follows from CB Lemma 2.3.14:

$$\lim_{n\to\infty} \left(1 + \frac{a_n}{n}\right)^n = e^a,$$

with $\lim_{n\to\infty} a_n = a$.

Remark:

- Read CB Example 2.3.13
- See CB Figure 2.3.3 for Poisson approximation to the binomial with $n = 15, p = 0.3$.
- *Question for you:* Draw (manually) a figure like CB Figure 2.3.3 for $n = 5, p = 0.3$.

Example (CB Example 3.2.5): A typesetter, on average, makes one error in every 500 words typeset. A typical page contains 300 words. What is the probability that there will be no more than two errors in five pages?

- $B(n,p) = B\left(1500, \frac{1}{500}\right)$:

$$\begin{aligned}\Pr(Y \le 2) &= \sum_{y=0}^{2} \binom{n}{y} p^y (1-p)^{n-y} \\ &= \sum_{y=0}^{2} \binom{1500}{y} \left(\frac{1}{500}\right)^y \left(1 - \frac{1}{500}\right)^{n-y} = 0.4230.\end{aligned}$$

- $P(\lambda) = P(np) = P(3)$:

$$\Pr(Y \le 2) \approx \sum_{y=0}^{2} e^{-\lambda} \frac{\lambda^y}{y!} = \sum_{y=0}^{2} e^{-3} \frac{3^y}{y!} = 0.4232.$$

2. Continuous Distributions

Uniform Distribution

A c.r.v. X follows a uniform probability distribution on $[a, b]$, denoted as $X \sim U[a, b]$, if its pdf

$$f_X(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & \text{otherwise.} \end{cases}$$

Because X is a bounded random variables, all moments exist. The kth moment

$$\begin{aligned} \mathbb{E}X^k &= \int_{-\infty}^{\infty} x^k f_X(x) dx \\ &= \frac{1}{b-a} \int_{-\infty}^{\infty} x^k dx \\ &= \frac{1}{b-a} \left. \frac{x^{k+1}}{k+1} \right|_a^b \\ &= \frac{1}{b-a} \frac{b^{k+1} - a^{k+1}}{k+1}. \end{aligned}$$

When $k = 1$,

$$\mu_X = \frac{1}{b-a}\frac{b^2-a^2}{2} = \frac{1}{2}(a+b).$$

When $k = 2$,

$$\mathbb{E}(X^2) = \frac{1}{b-a}\frac{b^3-a^3}{3} = \frac{1}{3}(a^2+b^2+ab),$$

where we have made the use of the formula $b^3 - a^3 = (b-a)(a^2+b^2+ab)$. It follows that the variance is

$$\sigma_X^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{1}{3}(a^2+b^2+ab) - \left[\frac{1}{2}(a+b)\right]^2 = \frac{1}{12}(b-a)^2.$$

The mgf is

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} e^{tx} f_X(x)dx \\ &= \int_a^b e^{tx} \frac{1}{b-a} dx \\ &= \frac{1}{t(b-a)} e^{tx} \Big|_a^b \\ &= \frac{1}{t(b-a)} \left(e^{tb} - e^{ta} \right). \end{aligned}$$

Question for you: Find μ_X and σ_X^2 using the mgf.

The Gamma Distribution

The gamma function:

$$\Gamma(\alpha) = \int_0^\infty z^{\alpha-1} e^{-z} dz, \qquad z \in [0, \infty),\ \alpha > 0, \tag{1}$$

has the following properties (CB Exercise 3.16):

$$\begin{aligned} \Gamma(\alpha+1) &= \alpha\Gamma(\alpha), \qquad \alpha > 0 \\ \Gamma(n) &= (n-1)! \text{ for } n \text{ a positive integer} \\ \Gamma\left(\frac{1}{2}\right) &= \sqrt{\pi} \\ \Gamma(1) &= 1 \end{aligned}$$

Gamma pdf:

- Note that

$$f_Z(z) := \frac{z^{\alpha-1}e^{-z}}{\Gamma(\alpha)}, \qquad z \in [0, \infty)$$

  is positive and $\int_0^\infty \frac{z^{\alpha-1}e^{-z}}{\Gamma(\alpha)}dz = 1$ from (1), and thus $f_Z(z)$ is a density of Z.

- Consider a monotonic transformation, $X = \beta Z$ with $\beta > 0$. Show that the density of X is

$$f_X(x) = \frac{x^{\alpha-1}e^{-x/\beta}}{\Gamma(\alpha)\beta^a}, \qquad 0 < x < \infty, \quad \alpha > 0, \ \beta > 0.$$

  The gamma distribution, denoted as $G(\alpha, \beta)$ with $\alpha, \beta > 0$ is a flexible family of distribution on $[0, \infty)$. Here, α is a shape parameter and β is a scale parameter controlling the spread of the distribution.

Remarks:

- $f_X\,(x) = f_Z(z)|\frac{dz}{dx}| = \frac{z^{\alpha-1}e^{-z}}{\Gamma(\alpha)}\frac{1}{\beta} = \frac{(x/\beta)^{\alpha-1}e^{-(x/\beta)}}{\Gamma(\alpha)}\frac{1}{\beta} = \frac{x^{\alpha-1}e^{-x/\beta}}{\Gamma(\alpha)\beta^a}$.
- $Z \sim G(\alpha, 1)$
- $X = \beta Z \sim G(\alpha, \beta)$
- Hence, $Y = \sigma X \sim G(\alpha, \sigma\beta)$ for $\sigma > 0$.

Gamma mgf:

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

$$\begin{aligned} M_X(t) &= \int_0^\infty e^{tx} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} dx \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty x^{\alpha-1} e^{-x(1-\beta t)/\beta} dx \\ &= \begin{cases} \left(\frac{1}{1-\beta t}\right)^\alpha & \text{if } t < \frac{1}{\beta} \\ \infty & \text{if } t \geq \frac{1}{\beta} \end{cases} \end{aligned}$$

Note that if $t \geq 1/\beta$, the mgf does not exist because $\int_0^\infty x^{\alpha-1} e^{-x(1-\beta t)/\beta} dx = \infty$.
See CB Example 2.3.8 for the derivation.

Gamma characteristic function:

The $G(\alpha, \beta)$ distribution has the characteristic function

$$\varphi(t) = \left(\frac{1}{1 - i\beta t}\right)^{\alpha}.$$

The χ^2_k distribution is a gamma $G(k/2, 2)$.

Remark: $\varphi(t)$ is bounded for all $-\infty < t < \infty$, while $M_X(t)$ is bounded for $t < \frac{1}{\beta}$.

Moments of gamma distribution:
The mean of $G(\alpha, \beta)$ distribution is

$$\frac{\varphi^{(1)}(0)}{i} = \left. \frac{-\alpha\left(1-\beta i t\right)^{-\alpha-1}(-\beta i)}{i} \right|_{t=0} = \alpha\beta$$

and the second moment is

$$\frac{\varphi^{(2)}(0)}{i^2} = \left. \frac{-\alpha(-\alpha-1)\left(1-\beta i t\right)^{-\alpha-2}(\beta i)^2}{i^2} \right|_{t=0} = \alpha(\alpha+1)\beta^2.$$

Hence the variance is

$$\alpha(\alpha+1)\beta^2 - (\alpha\beta)^2 = \alpha\beta^2.$$

See CB p. 100 for the derivation using the density.

Applications:

- The Gamma distribution has been used to model the distribution of the waiting time for economic events (e.g., unemployment duration, price duration, etc.).
- Cox, Ingersoll and Ross (CIR, 1985, *Econometrica*) proposes a diffusion model for the short-term interest rate r_t

$$dr_t = (\alpha + \beta r_t)dt + \sigma\sqrt{r_t}dW_t,$$

  where W_t is the standard Brownian motion. This model yields an analytic pricing formula for bond prices, and the marginal distribution of r_t is a *Gamma* distribution.

Exponential Distribution:

- The exponential distribution is a special case of the Gamma distribution $G(\alpha, \beta)$ with $\alpha = 1$ with its density $f_X(x) = \frac{1}{\beta} \exp\left(-\frac{x}{\beta}\right)$, $x > 0$.
- This is a gamma distribution with $\alpha = 1$, which is called an exponential distribution with density $h\left(y\right) = \beta^{-1} e^{-y/\beta}$, $y > 0$. We have shown that $\mathbb{E}\left(Y\right) = \beta$ and $Var(Y) = \beta^2$.
- Let X have uniform distribution with density $f(x) = 1$, $0 < x < 1$; zero, elsewhere. Find the density $h(y)$ of $Y = -\beta \ln X$, $\beta > 0$.

$$\begin{aligned} Y &= g(X) = -\beta \ln X \qquad 0 < X < 1 \\ g^{-1}(Y) &= X = \exp(-Y/\beta) \qquad 0 < Y < \infty \end{aligned}$$

$$\begin{aligned} H(y) &= \Pr(Y \leq y) = \Pr(g(X) \leq y) = \Pr(-\beta \ln X \leq y) \\ &= \Pr(\ln X \geq -y/\beta) = \Pr(X \geq \exp(-y/\beta)) \\ &= 1 - \Pr(X \leq \exp(-y/\beta)) \\ &= 1 - F(\exp(-y/\beta)) \end{aligned}$$

$$\begin{aligned} h(y) &= \frac{d}{dy} H(y) = \frac{d}{dy}\left[1 - F(\exp(-y/\beta))\right] \\ &= -f(\exp(-y/\beta)) \cdot \exp(-y/\beta) \cdot (-1/\beta) \\ &= \begin{cases} \frac{1}{\beta}\exp(-\frac{y}{\beta}) & 0 < y < \infty \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Application in labor economics:

Let T be the unemployment duration of a worker which has a pdf $f_T(t)$. The instantaneous probability that the worker will find a job after an unemployment duration of t, so-called **hazard rate,** is defined as

$$\begin{aligned}
\lambda(t) &= \lim_{\Delta t \to 0^+} \frac{\Pr(T \le t + \Delta t \mid T \ge t)}{\Delta t} \\
&= \lim_{\Delta t \to 0^+} \frac{1}{\Delta t} \frac{\Pr(t \le T \le t + \Delta t)}{\Pr(T \ge t)} \\
&= \left[\lim_{\Delta t \to 0^+} \frac{\int_t^{t+\Delta t} f_T(u)du}{\Delta t} \right] \cdot \frac{1}{\Pr(T \ge t)} \\
&= \frac{f_T(t)}{\Pr(T \ge t)}.
\end{aligned}$$

A simplest example is to assume that

$$f_T(t) = \frac{1}{\beta} \exp\left(-\frac{t}{\beta}\right), \quad t > 0.$$

Then it can be shown that the hazard rate will become

$$\begin{aligned}
\lambda(t) &= \frac{f_T(t)}{\Pr(T \geq t)} \\
&= \frac{\frac{1}{\beta} \exp\left(-\frac{t}{\beta}\right)}{\exp\left(-\frac{t}{\beta}\right)} \\
&= \frac{1}{\beta},
\end{aligned}$$

which is constant over time.

Applications in financial economics:

- The autoregressive conditional duration (ACD) model for financial events. See Engle and Russell (1996, *Econometrica*).
- In financial econometrics, an empirical stylized fact of high-frequency financial return X_t is that $|X_t|$ approximately follows an exponential distribution (see Ding, Granger and Engle 1993, *Journal of Empirical Finance*). Granger and Ding (1995) established a set of temporal and distributional properties that the returns are well characterized by the double exponential distribution.

Double Exponential Distribution

The double exponential distribution is formed by reflecting the exponential distribution into a mirror around the mean. The pdf is given by

$$
\begin{aligned}
f_X(x|\mu,\sigma) &= \frac{1}{2\sigma}\exp\left(-\frac{|x-\mu|}{\sigma}\right), \\
-\infty &< x < \infty,\ -\infty < \mu < \infty,\ \sigma > 0.
\end{aligned}
$$

- The double exponential distribution has a fatter tail than the normal distribution. But it still retains all of its moments.
- It has a peak at $x = \mu$ (derivatives does not exist).
- The double exponential distribution is also called the Laplace distribution.

Exercise: Show that $E(X) = \mu$ and $Var(X) = 2\sigma^2$. (Hint. See CB pp. 110-111. CB Exercise 3.22)

The Normal Distribution

The **standard normal** distributions, denoted by $\mathcal{N}\left(0,1\right)$ is a continuous distribution with density

$$f_X\left(x\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \qquad x \in \mathbb{R}.$$

Question for you: Show that $f_X\left(x\right)$ is a pdf, i.e., $f_X\left(x\right) > 0$ and

$$\int_{-\infty}^{\infty} f_X\left(x\right) dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = 1.$$

Proof:

$$
\begin{aligned}
\left[\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx\right]^2 &= \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2+y^2)} dxdy \\
&= \int_{0}^{\infty}\int_{0}^{2\pi} e^{-\frac{1}{2}r^2} r d\theta dr \\
&= \left(\int_{0}^{2\pi} d\theta\right)\int_{0}^{\infty} e^{-\frac{1}{2}r^2} r dr \\
&= 2\pi \int_{0}^{\infty} e^{-\frac{1}{2}r^2} r dr \\
&= 2\pi \left[-e^{r^2/2}\Big|_0^{\infty}\right] \\
&= 2\pi.
\end{aligned}
\tag{2}
$$

Remarks on Eq (2):

- Set $x = r\cos(\theta)$ and $y = r\sin(\theta)$.
- Why $dx\ dy = r\ dr\ d\theta$ in (2)? In other words, why is r there? It is the Jacobian of the bivariate (polar) transformation $g : (x, y) \to (r, \theta)$ or $(x, y) = g^{-1}(r, \theta)$. Set $x = r\cos(\theta) = g_1^{-1}(r, \theta)$ and $y = r\sin(\theta) = g_2^{-1}(r, \theta)$.
- See CB Chapter 4, where we will discuss that r is the Jacobian of the bivariate transformation $g : (x, y) \to (r, \theta)$.
- See the next page, or you can wait till CB Section 4.3.
- See CB pp. 103-104.

Bivariate nonlinear transformation (CB Section 4.3)

Consider $Y = g(X)$, where $g(\cdot)$ is a one-to-one mapping from $\mathbb{R}^k$ to $\mathbb{R}^k$ with differentiable inverse $X = g^{-1}(Y)$. Then the distribution function of Y has density

$$h_Y(y) = f_X\left(g^{-1}(y)\right)\left|\det(\partial g^{-1}(y)/\partial y)\right|.$$

where $\det(\partial g^{-1}(y)/\partial y)$ is called the Jacobian of the transformation, and $\partial g^{-1}(y)/\partial y$ is defined as

$$\frac{\partial g^{-1}(y)}{\partial y} = \begin{bmatrix} \frac{\partial g_1^{-1}(y_1,\cdots,y_k)}{\partial y_1} & \cdots & \frac{\partial g_1^{-1}(y_1,\cdots,y_k)}{\partial y_k} \\ \vdots & & \vdots \\ \frac{\partial g_k^{-1}(y_1,\cdots,y_k)}{\partial y_1} & \cdots & \frac{\partial g_k^{-1}(y_1,\cdots,y_k)}{\partial y_k} \end{bmatrix},$$

where $g_i^{-1}(y_1,\cdots,y_k)$ is the i-th component of $g^{-1}(y_1,\cdots,y_k)$.

The characteristic function of the standard normal distribution

$$
\begin{aligned}
\varphi_X(t) &= \int_{-\infty}^{\infty} e^{itx} dF(x) \\
&= \int_{-\infty}^{\infty} e^{itx} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \\
&= \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-it)^2} dx \right) \cdot \left(e^{-\frac{1}{2}t^2} \right) \\
&= \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \right) \cdot \left(e^{-\frac{1}{2}t^2} \right) \\
&= e^{-\frac{1}{2}t^2}.
\end{aligned}
$$

Question for you: Similarly, we can derive the mgf of the standard normal distribution.

Let X be distributed as $\mathcal{N}(0,1)$ and let

$$Y = \sigma X + \mu, \qquad \sigma > 0.$$

Then

$$\begin{aligned} H(y) &= \Pr(Y \leq y) = \Pr(\sigma X + \mu \leq y) \\ &= \Pr\left(X \leq \frac{y-\mu}{\sigma}\right) = F\left(\frac{y-\mu}{\sigma}\right), \end{aligned}$$

$$\begin{aligned} h(y) &= \frac{d}{dy} H(y) = \frac{d}{dy} F\left(\frac{y-\mu}{\sigma}\right) \\ &= f\left(\frac{y-\mu}{\sigma}\right) \cdot \frac{1}{\sigma} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}, y \in \mathbb{R}. \end{aligned}$$

pdf of normal distribution $\mathcal{N}\left(\mu, \sigma^2\right)$:

$$h\left(y\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}, \qquad y \in \mathbb{R}. \tag{3}$$

Moments:

$$\begin{aligned} \mathbb{E}Y &= \sigma\mathbb{E}X + \mu = \mu, \\ \mathbb{E}Y^2 &= \sigma^2\mathbb{E}X^2 + 2\sigma\mu\mathbb{E}X + \mu^2 = \sigma^2 + \mu^2, \\ Var(Y) &= \mathbb{E}Y^2 - \left(\mathbb{E}Y\right)^2 = \sigma^2. \end{aligned}$$

The characteristic function:

$$\begin{aligned}\varphi_Y(t) &= \mathbb{E}e^{itY} \\ &= \mathbb{E}e^{it(\sigma X+\mu)} \\ &= \left(\mathbb{E}e^{it\sigma X}\right)\cdot e^{it\mu} \\ &= \varphi_X(t\sigma)\cdot e^{it\mu} \\ &= e^{-\frac{1}{2}t^2\sigma^2}\cdot e^{it\mu} \\ &= e^{it\mu-\frac{1}{2}t^2\sigma^2}.\end{aligned}$$

Summarizing, we have:

Theorem. The normal distribution $\mathcal{N}\left(\mu,\sigma^2\right)$ with the density (3) has mean μ, variance σ^2, and characteristic function $\varphi(t)=\exp\left(it\mu-\frac{1}{2}t^2\sigma^2\right).$ □

Theorem: If $X_1, \cdots, X_n$ are independent random variables with $\mathcal{N}\left(\mu_i, \sigma_i^2\right)$ distributions for $i = 1, \cdots, n$, then

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

$\square$

Remark. If X is distributed $\mathcal{N}(\mu_x, \sigma_X^2)$ and Y is distributed $\mathcal{N}(\mu_x, \sigma_X^2)$, and X and Y are independent then

$$\begin{aligned} X + Y &\sim \mathcal{N}\left(\mu_X + \mu_Y,\ \sigma_X^2 + \sigma_Y^2\right), \\ X - Y &\sim \mathcal{N}\left(\mu_X - \mu_Y,\ \sigma_X^2 + \sigma_Y^2\right). \end{aligned}$$

Remarks: The normal distribution arises in a natural way *for the sample statistics* from the central limit theorem (CLT). We will see the following in Chapter 5.

Theorem [Lindberg-Levy]: Let $X_1, \cdots, X_n$ be a sequence of independently identically distributed (i.i.d.) random variables with mean μ and finite variance σ^2. Let $\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i$. Then

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} Y,$$

where Y is distributed as $\mathcal{N}(0,1)$. $\square$

Remark. $\sqrt{n}\left(\bar{X}_n - \mu\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$.

Moments of the normal distribution:

Let $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ with $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$.

- *Question for you:* Show that the mean $\mathbb{E}(X) = \mu$.

$$
\begin{aligned}
\mathbb{E}(X) &= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx \\
&= \int_{-\infty}^{\infty} (x - \mu + \mu) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx \\
&= \int_{-\infty}^{\infty} (x - \mu) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx + \mu \\
&= \int_{-\infty}^{\infty} y \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}y^2} dy + \mu \\
&= 0 + \mu
\end{aligned}
$$

because $g(y) = y \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}y^2} = -g(-y)$ is an odd function.

- *Question for you:* Show that the variance $Var(X) = \sigma^2$.

$$\begin{aligned}
Var(X) &= \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx \\
&= \int_{-\infty}^{\infty} y^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}} dy \\
&= -\sigma^2 \int_{-\infty}^{\infty} y d\left[\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}}\right] \\
&= \int u dv \\
&= uv - \int v du \\
&= -\sigma^2 \left[y \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}} \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}} dy \right] \\
&= -\sigma^2 [0 - 1] \\
&= \sigma^2.
\end{aligned}$$

Question for you: Show that $\mathbb{E}(X) = \mu$ and $Var(X) = \sigma^2$ using the mgf.

Normal mgf:

$$M_X(t) = E(e^{tX}) = e^{t\mu + \frac{1}{2}t^2\sigma^2}.$$

Show that:

$$\begin{aligned} M_X^{(1)}(t)|_{t=0} &= \mu \\ M_X^{(2)}(t)|_{t=0} &= \sigma^2 + \mu^2. \end{aligned}$$

Higher moments:
All centered odd moments $\mathbb{E}\left[(X-\mu)^{2n+1}\right] = 0$ for all integers $n \geq 0$,because the normal distribution is symmetric around μ. All centered even moments, for all integers $n > 0$, are

$$\mathbb{E}\left[(X-\mu)^{2n}\right] = \frac{1}{\pi^{1/2}}\Gamma(n+\frac{1}{2})2^n\sigma^{2n}.$$

A special case when $n = 2$:

$$\mathbb{E}\left[(X-\mu)^{4}\right] = \frac{1}{\pi^{1/2}}\Gamma(2+\frac{1}{2})4\sigma^4 = \frac{1}{\pi^{1/2}}\Gamma(2+\frac{1}{2})4\sigma^4 = 3\sigma^4.$$

Normal approximation of the binomial distribution:
Read: CB Example 3.3.2, CB Figure 3.3.2
Question for you: Compare the following two distributions by simulation.

- $X \sim$ Binomial(n, p) distribution with $n = 25$, $p = 0.6$.
  - Explain how you can generate $X \sim$ Binomial(n, p) distribution.
  - Generate 1000 observations from Binomial(n, p) distribution.
  - Compute the sample mean and the sample variance of your random sample. Draw the histogram.
- $Y \sim \mathcal{N}(np, np(1-p))$ with $n = 25$, $p = 0.6$.
  - Generate 1000 observations from $\mathcal{N}(np, np(1-p))$ distribution.
  - Compute the sample mean and the sample variance of your random sample. Draw the histogram.

Log-normal Distribution

The pdf of the lognormal distribution:
A c.r.v. X follows a lognormal distribution if its pdf

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\frac{1}{x}\exp\left[-\frac{1}{2\sigma^2}(\ln x - \mu)^2\right], \qquad x > 0.$$

It fact, $X = \exp(Y)$, where $Y \sim \mathcal{N}\left(\mu, \sigma^2\right)$. X is called a lognormal random variable because its logarithm follows a normal distribution.

The mgf of the lognormal distribution:
Recall the MGF of a normal r.v. Y is

$$M_Y(t) = \mathbb{E}(e^{tY}) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right).$$

It follows that all moments of X exist:

$$\mathbb{E}(X^k) = \mathbb{E}(e^{kY}) = M_Y(k) = \exp\left(k\mu + \frac{1}{2}k^2\sigma^2\right), \qquad k = 1, 2, \ldots$$

In particular,

$$\begin{aligned} \mu_X &= \mathbb{E}(e^Y) = M_Y(1) = \exp\left(\mu + \frac{1}{2}\sigma^2\right), \\ \sigma_X^2 &= \mathbb{E}(X^2) - \mu_X^2 = \exp(2\mu)\left[\exp\left(2\sigma^2\right) - \exp\left(\sigma^2\right)\right]. \end{aligned}$$

Although all moments exist, the MGF does not exist for a lognormal distribution. Why?

$$
\begin{aligned}
M_X(t) &= \mathbb{E}(e^{tX}) \\
&= \int_0^\infty e^{tx} \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{x} \exp\left[-\frac{1}{2\sigma^2} (\ln x - \mu)^2\right] dx \\
&= \int_0^\infty \exp(t \exp(\ln x)) \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2} (\ln x - \mu)^2\right] d \ln x \\
&= \int_0^\infty \exp(t \exp(y)) \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2} (y - \mu)^2\right] dy
\end{aligned}
$$

$$
\begin{aligned}
&= \int_0^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[t\exp(y) - \frac{1}{2\sigma^2}\left(y-\mu\right)^2\right] dy \\
&\geq \int_c^{c+1} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[t\exp(y) - \frac{1}{2\sigma^2}\left(y-\mu\right)^2\right] dy \qquad \text{for any } c > 0 \\
&\geq \frac{1}{\sqrt{2\pi}\sigma} \exp\left[t\exp(c) - \frac{1}{2\sigma^2}\left(c-\mu\right)^2\right] (c+1-c) \\
&= \frac{1}{\sqrt{2\pi}\sigma} \exp\left[t\exp(c) - \frac{1}{2\sigma^2}\left(c-\mu\right)^2\right] \rightarrow \infty \text{ as } c \rightarrow \infty
\end{aligned}
$$

because $\left[t\exp(c) - \frac{1}{2\sigma^2}\left(c-\mu\right)^2\right] \rightarrow \infty$ as $c \rightarrow \infty$ for $t > 0$.

Question for you: Discuss the usefulness of the logarithmic transformation of an economic variable.

- The lognormal distribution is similar in appearance to the Gamma distribution. See CB Figure 3.3.6.
- The lognormal distribution is very popular in modelling applications when the variable of interest is skewed to the right.
- Incomes are necessarily skewed to the right, and modelling with a lognormal allows the use of normal-theory statistics on **log(income)**, a very convenient circumstance.

Example: The lognormal distribution and the Black-Scholes formula for option pricing. If X is lognormal(μ, σ^2), then $\ln X \sim \mathcal{N}(\mu, \sigma^2)$. Let $X_t = P_t / P_{t-1}$.

$$\ln X_t = \ln(P_t/P_{t-1}) = \ln P_t - \ln P_{t-1} = \frac{P_t - P_{t-1}}{P_t} \text{ is } \mathcal{N}(\mu, \sigma^2).$$

The lognormal distribution is very popular in finance. A very important process in economics and finance is a geometric random walk:

$$d \ln X_t = \mu dt + \sigma dW_t,$$

where W_t is a Brownian motion. Then

$$\begin{aligned} \ln X_t &= \mu t + \sigma \int_0^t dW_s \\ X_t &= \exp\left(\mu t + \sigma \int_0^t dW_s\right) \end{aligned}$$

is a lognormal process.

Cauchy Distribution

- The Cauchy distribution has the pdf
$$f_X(x) = \frac{1}{\pi}\frac{1}{1+(x-\theta)^2}, \qquad -\infty < x < \infty, \ -\infty < \theta < \infty.$$
- The pdf is integrated to one as it should: $\int_{-\infty}^{\infty} \frac{1}{\pi}\frac{1}{1+(x-\theta)^2}dx = 1$ (CB p. 108)
- CB Figure 3.3.5. The pdf is symmetric, bell-shaped on $(-\infty, \infty)$, with thicker tails than the standard normal density.
- CB Example 2.2.4. No moments of the Cauchy distribution exist because
$$\mathbb{E}|X| = \int_{-\infty}^{\infty} \frac{1}{\pi}\frac{|x|}{1+(x-\theta)^2}dx = \infty$$
- The mgf does not exist. (Why?)
- The median of the Cauchy distribution exist and it is θ. CB Exercise 2.17(b).
- The ratio of two independent standard normals has a Cauchy distribution. CB Example 4.3.6.

The Chi-squared Distribution

Let $X_1, \cdots, X_k$ be independent $\mathcal{N}(0, 1)$ distributed random variables and let $Y_k = \sum_{j=1}^{k} X_j^2$. Then Y_k is distributed as χ_k^2 with k degrees of freedom.

The pdf of the χ_1^2 distribution:
Let us consider the case $k = 1 : Y = X^2$, where $X \sim \mathcal{N}(0, 1)$.

$$\begin{aligned} H(y) &= \Pr(Y \leq y) \\ &= \Pr(X^2 \leq y) \\ &= \Pr\left(-\sqrt{y} \leq X \leq \sqrt{y}\right) \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \\ &= 2\int_0^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx, \qquad y \geq 0. \end{aligned}$$

If we change the variable of integration by writing $x = \sqrt{u}$, then $dx = \frac{1}{2\sqrt{u}} du$, and

$$
\begin{aligned}
H(y) &= 2\int_0^y \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u} \frac{1}{2\sqrt{u}} du \\
&= \int_0^y \frac{1}{\sqrt{2\pi}\sqrt{u}} e^{-\frac{1}{2}u} du, \qquad y \geq 0,
\end{aligned}
$$

where the integral is taken over x from $x = 0 = u$ to $x = \sqrt{y} = \sqrt{u}$ (i.e., over u from $u = 0$ to $u = y$). Hence, the density of Y is

$$
h(y) = \frac{1}{\sqrt{2\pi}\sqrt{y}} e^{-\frac{1}{2}y}, \qquad y > 0.
$$

Remark: Recall that we have obtained this in Lecture 2. This tells that there are more than one ways to obtain the density of transformed a random variable.

The characteristic function of the χ_1^2 distribution:

$$
\begin{aligned}
\varphi_1(t) &= \mathbb{E}e^{ity} = \int_0^\infty e^{ity} h(y)dy \\
&= \int_0^\infty e^{ity} \frac{1}{\sqrt{2\pi}\sqrt{y}} e^{-\frac{1}{2}y} dy \\
&= \int_0^\infty \frac{e^{-\frac{1}{2}(1-2it)y}}{\sqrt{2\pi}\sqrt{y}} dy \\
&= \int_0^\infty \frac{e^{-\frac{1}{2}(1-2it)y}}{\sqrt{2\pi}\sqrt{y}\sqrt{(1-2it)}} d[(1-2it)y] \cdot \frac{1}{\sqrt{(1-2it)}} \\
&= \int_0^\infty \frac{e^{-\frac{1}{2}u}}{\sqrt{2\pi}\sqrt{u}} du \cdot \frac{1}{\sqrt{(1-2it)}} \\
&= \frac{1}{\sqrt{(1-2it)}} = (1-2it)^{-\frac{1}{2}}
\end{aligned}
$$

where we let $u \equiv (1-2it)y$, which gives $u/(1-2it) = y$, $du = (1-2it)dy$, and $du/(1-2it) = dy$.

The characteristic function of the χ_k^2 distribution:

$$\begin{aligned}
\varphi_k(t) &= \mathbb{E}e^{itY} \\
&= \mathbb{E}e^{it\sum_{j=1}^{k} X_j^2} \\
&= \mathbb{E}\prod_{j=1}^{k} e^{itX_j^2} \\
&= \prod_{j=1}^{k} \mathbb{E}e^{itX_j^2} \\
&= \prod_{j=1}^{k} \varphi_j(t) \\
&= (1-2it)^{-\frac{1}{2}k}.
\end{aligned}$$

Question for you: Therefore, the χ_k^2 distribution is a special case of the gamma distribution $G(\alpha, \beta)$, with $\alpha =$ and $\beta =$.

Moments of the χ_k^2 distribution:

$$\begin{aligned} \mathbb{E}Y_k &= \mathbb{E}\sum_{j=1}^{k} X_j^2 = \sum_{j=1}^{k} \mathbb{E}X_j^2 = \sum_{j=1}^{k} 1 = k, \\ Var(Y_k) &= \sum_{j=1}^{k} Var(X_j^2) = \sum_{j=1}^{k} \left[\mathbb{E}(X_j^4) - \mathbb{E}(X_j^2)^2\right] \\ &= \sum_{j=1}^{k} \left[\mathbb{E}(X_j^4) - 1\right] = \sum_{j=1}^{k} [3-1] = 2k. \end{aligned}$$

Exercise. Show that $\mathbb{E}(X_j^4) = 3$.

Summarizing, we have

Theorem. If Y is χ_k^2 distributed, then it has mean k, variance $2k$, and characteristic function $\varphi_Y(t) = (1-2it)^{-k/2}$. $\square$

Theorem: Let $X \sim \chi^2_m$ and $Y \sim \chi^2_n$. If X and Y are independent, then $X + Y \sim \chi^2_{m+n}$. $\square$

Proof.

$$\begin{aligned}
\varphi_X(t) &= \mathbb{E}e^{itX} = (1-2it)^{-\frac{1}{2}m} \\
\varphi_Y(t) &= \mathbb{E}e^{itY} = (1-2it)^{-\frac{1}{2}n} \\
\varphi_{X+Y}(t) &= \mathbb{E}e^{it(X+Y)} = \mathbb{E}e^{itX} \cdot \mathbb{E}e^{itY} = (1-2it)^{-\frac{1}{2}(m+n)}.
\end{aligned}$$

Exercise. Find sequences $(a_n), (b_n)$ such that

$$\frac{\chi_n^2 - a_n}{b_n} \xrightarrow{d} \mathcal{N}(0,1) \text{ as } n \to \infty.$$

Solution. $a_n = n$ and $b_n = \sqrt{2n}$, by CLT.

Remark: The Chi-square distribution is one of the most important distributions in statistics and econometrics. Many popular test statistics in statistics and econometrics have an asymptotic Chi-square distribution, as we study later.

Beta Distribution

The Beta distribution has a pdf

$$f_X(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}, \qquad 0 < x < 1,$$

where $\alpha > 0, \beta > 0$, and

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

It can be shown that

$$\begin{aligned} \mu_X &= \frac{\alpha}{\alpha+\beta}, \\ \sigma_X^2 &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}. \end{aligned}$$

See CB pp. 106-107.

Remarks:

- Beta(α, β) is one of the few well-known distributions that have bounded support.
- The standard uniform distribution on the unit interval [0, 1] is a special case of Beta[1, 1] with $\alpha = \beta = 1$.
- Because the support of a Beta(α, β) is [0, 1], the Beta distribution can be used to model **proportions**. See CB Example 4.4.6. You can read this example later when we study the law of iterated expectation (CB Theorem 4.4.3.).

- Granger (1980) uses the Beta distribution for the marginal propensities to consume for individual consumers and show that the sum of "short memory" time series processes can display a "long memory" property.
- The shape of the Beta(α, β) distribution depends on the values of parameters (α, β). See CB Figure 3.3.3.
- When $\alpha = \beta$, Beta(α, α) has a symmetric density with mean $\mu_X = \frac{1}{2}$. CB Figure 3.3.4.
- Ghysel et al (2006) use the Beta(α, β) distribution for a parsimonious distributed lag model, for their model called MIDAS (mixed data sampling) model.

The Student-t Distribution

Let $Z \sim \mathcal{N}(0,1)$ and let $Y \sim \chi^2_k$. If Z and Y are independent, then

$$X = \frac{Z}{\sqrt{Y/k}}$$

is distributed as Student t with k degrees of freedom, denoted by t_k.

Remark: The mean of the t_k distribution does not exist for $k = 1$ and is equal to zero for $k \geq 2$. The second moment of t_k is infinite for $k = 1, 2$. For $k \geq 3$, the variance of the t_k distribution is $k/(k-2)$. For $k = 1$, the t_k distribution is the Cauchy distribution.

Remark: For $k \to \infty$, $t_k \xrightarrow{d} \mathcal{N}(0,1)$. Why?
$Y/k = \frac{1}{k}\sum_{j=1}^{k} X_j^2 \xrightarrow{p} \mathbb{E}X^2 = 1$ by WLLN, where $X_j,\ j = 1, \ldots, k \sim i.i.d. \mathcal{N}(0,1)$. As $\sqrt{Y/k} \xrightarrow{p} 1$ as $\sqrt{\cdot}$ is continuous. Thus $X = \frac{Z}{\sqrt{Y/k}} \xrightarrow{d} \mathcal{N}(0,1)$.

The Fisher's F Distribution

Let $U \sim \chi^2_m$ and let $V \sim \chi^2_n$. If U and V are independent, then $X = \frac{U/m}{V/n}$ is distributed as $F\left(m, n\right)$.

Remark: $t_k^2 \equiv F\left(1, k\right)$. $F\left(m, n\right) = 1/F\left(n, m\right)$.

Remark: For $n \to \infty$, $mF\left(m, n\right) \xrightarrow{d} \chi^2_m$. Why? Because $mX = \frac{U}{V/n} \xrightarrow{d} U \sim \chi^2_m$ and $V/n \xrightarrow{p} 1$ as $n \to \infty$.

Exponential Families

Definition: A family of pdfs or pmfs is called an exponential family if it can be expressed as

$$f_X(x|\theta) = h(x)c(\theta)\exp\left(\sum_{i=1}^{k} w_i(\theta)t_i(x)\right),$$

where $h(x) \geq 0$ and $c(\theta) \geq 0$. $\square$

Binomial Exponential Family

$$
\begin{aligned}
f_X(x) &= \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, 2, \ldots, n \\
&= \binom{n}{x} \left(\frac{p}{1-p}\right)^x (1-p)^n \\
&= \binom{n}{x} (1-p)^n \exp\left(\log\left(\frac{p}{1-p}\right) \cdot x\right) \\
&= h(x) c(\theta) \exp\left(\sum_{i=1}^{k} w_i(\theta) t_i(x)\right)
\end{aligned}
$$

with

$k = 1, \theta = p, h(x) = \binom{n}{x}, c(\theta) = (1-p)^n, w_1(p) = \log\left(\frac{p}{1-p}\right),$

and $t_1(x) = x$.

Normal Exponential Family

$$
\begin{aligned}
f_X(x) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) \\
&= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \exp\left(-\frac{x^2}{2\sigma^2} + \frac{\mu x}{\sigma^2}\right) \\
&= \left[\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\mu^2}{2\sigma^2}\right)\right] \exp\left(\frac{1}{\sigma^2}\left(-\frac{x^2}{2}\right) + \left(\frac{\mu}{\sigma^2}\right)x\right) \\
&= h(x)c(\theta)\exp\left(\sum_{i=1}^{k} w_i(\theta)t_i(x)\right)
\end{aligned}
$$

with $k = 2, \theta = (\mu, \sigma^2), h(x) = 1, c(\theta) = \left[\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\mu^2}{2\sigma^2}\right)\right], \ldots$

Remarks:

- White (1982, *Econometrica*) shows that the QMLE is consistent and asymptotically normal when the pdf or pmf belongs to the normal exponential family.
- Gourieroux, Monfort, and Trognon (1984, *Econometrica*) show that the QMLE is consistent and asymptotically normal when the pdf or pmf belongs to an exponential family.
- See White (1994 *Estimation, Inference, and Specification Analysis*, Cambridge Univerity Press) for a complete analysis on consistency and asymptotic normality of QMLE.
- Other examples in the exponential family: Bernoulli, binomial, Poisson, negative binomial, hypergeometric, and logarithmic series densities (e.g., Johanson and Katz 1970) and their multivariate generalizations such as the multinomial and negatove multinomial, as well as gamma (including exponential, χ^2), Pareto, and Weibull distributions.