MIT 18.642

Probability Theory

Dr. Kempthorne

Fall 2024

Terminology and Review

Random Variables: Discrete, Continuous, Mixed

- Discrete Random Variable (outcomes countable)
    - Counter-party default (1=default/0=no default)
    - FOMC Decision on Fed Funds Rate
    - Indicator of Black-Swan event within next 3 months
    - Side ("buy" or "sell") of next market order AAPL stock
    - Share-size of next market order for AAPL stock
- Continuous Random Variable
    - Asset value (stock, currency, future, bond, ...)
    - Waiting time to next market order for AAPL stock
- Mixed (continuous and discrete) Variable
    - Value of a stock in 6 months which may go
      bankrupt (Value=0)
      private (Value = net buyout price)

Terminology and Review

Probability model for a random variable X

- Sample space of X
  $$\mathcal{X} = \{\text{all possible outcomes } X = x\}$$
- Probability mass function for discrete X: $f_X(x)$
  $$f_X(x) = P(X = x), \text{ for all } x \in \mathcal{X}$$
  $$\sum_{x \in \mathcal{X}} f_X(x) = \sum_{x \in \mathcal{X}} P(X = x) = 1$$
- Probability density function for continuous X: $f_X(x)$
  When $\mathcal{X} \subset R = (-\infty, +\infty)$.
  $$P(X \in [x, x + dx]) = f_X(x)dx$$
  $$\int_{x \in \mathcal{X}} f_X(x)dx = \int_{-\infty}^{+\infty} f_X(x)dx = 1$$

Cumulative Distribution Function

$$F_X(x) = P(X \leq x),\ x \in \mathcal{X} \subset R.$$

- For discrete X:
$$F_X(x) = \sum_{x' \leq x} f_X(x')$$
- For continuous X:
$$F_X(x) = \int_{-\infty}^{x} f_X(u)du$$

Event and Event Probability : $A \subset \mathcal{X}$, $P(A) = P(X \in A)$

- For discrete X:
$$P(A) = P(X \in A) = \sum_{x \in A} f_X(x)$$
- For continuous X:
$$P(A) = P(X \in A) = \int_{x \in A} f_X(x)dx$$

Expectations/Moments/Skewness/Kurtosis

- The **expectation**/mean/first-moment of random variable X

$$\mu = E[X] = \begin{cases} \int_{\mathcal{X}} x f_X(x) dx & , if \quad X \ continuous \\ \sum_x x f_X(x) & , if \quad X \ discrete \end{cases}$$

- The **k-th moment** of random variable X $(k = 1, 2, \ldots)$

$$m_k = E[X^k] = \begin{cases} \int_{\mathcal{X}} x^k f_X(x) dx & , if \quad X \ continuous \\ \sum_x x^k f_X(x) & , if \quad X \ discrete \end{cases}$$

- The **variance** of a random variable

$$var(X) = E([X - E(X)]^2) = E(X^2) - [E(X)]^2 = m_2 - m_1^2$$

Terminology and Review

Standard Deviation

$$\sigma = \sqrt{var(X)} = \sqrt{m_2 - m_1^2} \qquad \text{same (!) units as } X$$

Skewness

$$\gamma = E[(X-\mu)^3]/\sigma^3 = E[(\tfrac{X-\mu}{\sigma})^3]. \text{ (no units!)}$$

- $\gamma = 0$: X is symmetric about μ
- $\gamma > 0$: X has positive skew (Long right tail)
  high probability of large positive values
- $\gamma < 0$: X has negative skew (Long left tail)
  high probability of large negative values

Kurtosis

$$\kappa = E[(X-\mu)^4]/\sigma^4 = E[(\tfrac{X-\mu}{\sigma})^4] \text{ (no units!)}$$

$\kappa > 3 \Longleftrightarrow$ **fat-tailed** (relative to Gaussian)

Normal/Gaussian Distribution

Definition. A **Normal (Gaussian)** random variable $X \sim N(\mu, \sigma^2)$ has density function:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}, \quad -\infty < x < +\infty.$$

with mean and variance parameters:

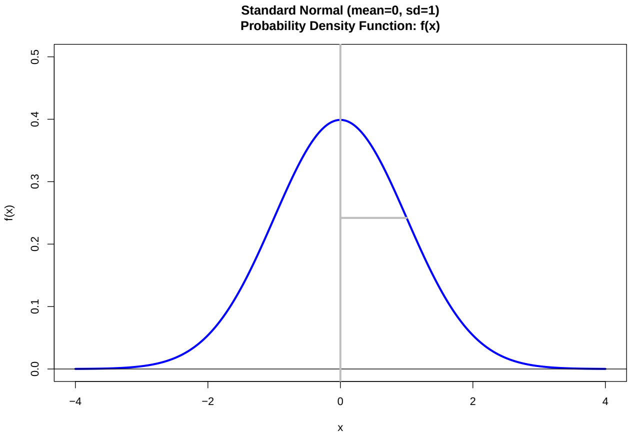
$$\begin{array}{rcccl} \mu & = & E[X] & = & \int_{-\infty}^{+\infty} x f(x) dx \\ \sigma^2 & = & E[(X-\mu)^2] & = & \int_{-\infty}^{+\infty} (x-\mu)^2 f(x) dx \end{array}$$

Note: $-\infty < \mu < +\infty$, and $\sigma^2 > 0$.

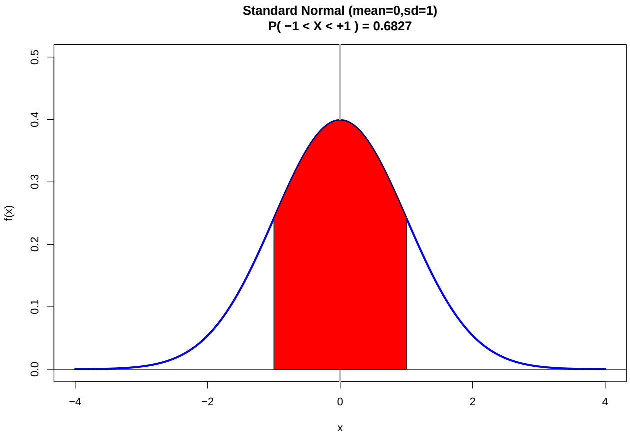
Properties:

- Density function is symmetric about $x = \mu$.
  $f(\mu + x^*) = f(\mu - x^*)$. (zero skewness $\gamma = 0$)
- $f(x)$ is a maximum at $x = \mu$.
- $f''(x) = 0$ at $x = \mu + \sigma$ and $x = \mu - \sigma$
  (inflection points of bell curve)

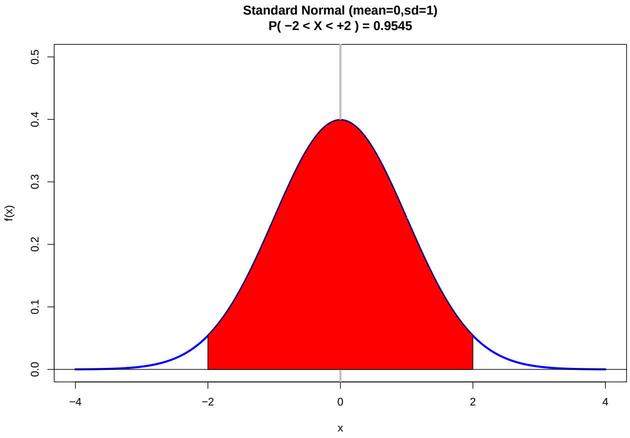
Standard Normal Model



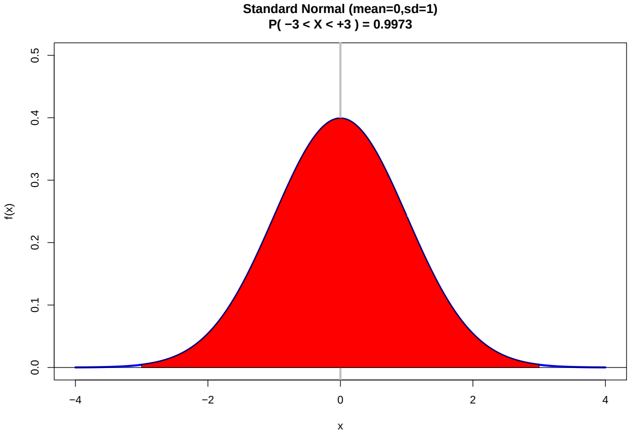
Standard Normal Model



Standard Normal Model



Standard Normal Model



Lognormal Distribution

Definition

- Random variable Y has the *lognormal*(μ, σ^2) distribution if
  $x = log(Y)$ has a $N(\mu, \sigma^2)$ distribution.
- Equivalently, suppose random variable X has distribution
  $N(\mu, \sigma^2)$ with mean μ and variance σ^2.
  Define random variable Y by transforming X
  $Y = e^X$
  Then the distribution of Y is *lognormal*(μ, σ^2).

Change-of-Variables Theorem

Transforming a Random Variable
Suppose that X is a continuous real-valued random variable.

- Let $g : R \to R$ be a continuous, differentiable monotone increasing function, and define the random variable Y as
  $$Y = g(X)$$
- Let $h : R \to R$ denote the inverse function of g, also continuous, with derivative $h'(y) = \frac{d}{dy}h(y)$
  $$X = h(Y)$$
  E.g., $y = g(x) = e^x$ and $h(y) = ln(y) = g^{-1}(y) = x$

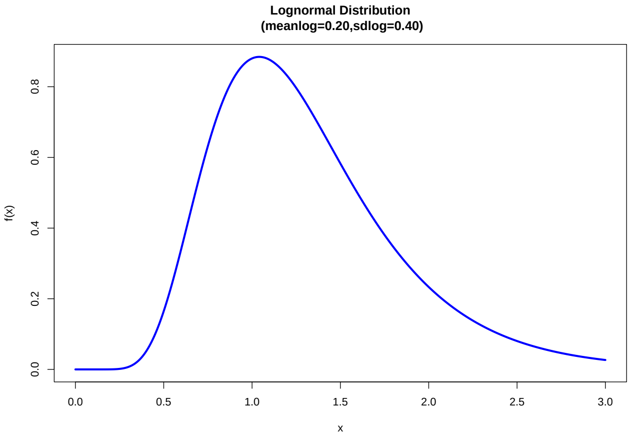
Theorem (Change-of-Variables). If $F_X(x)$ is the cumulative distribution function of X then
$$F_Y(y) = F_X(h(y))$$
If $f_X(x)$ is the probability density function of X then
$$f_Y(y) = f_x(h(y))h'(y)$$

Lognormal Distribution Let Y be a *lognormal*(μ, σ^2) distribution. Then $X = \ln Y$ is *Normal*(μ, σ^2)

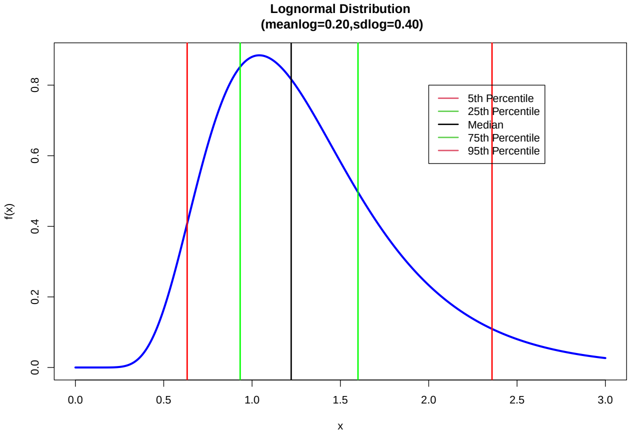
- The probability density function of Y is

$$
\begin{aligned}
f_Y(y) &= \tfrac{d}{dy} P(Y \le y) = \tfrac{d}{dy} P(\ln Y \le \ln y) \\
&= [f_X(\ln y)] \frac{d}{dy}(\ln y) \\
&= \left[\frac{1}{\sigma\sqrt{2\pi}} e^{-\dfrac{(\ln y - \mu)^2}{2\sigma^2}} \right] \times \frac{1}{y}, \quad y > 0
\end{aligned}
$$

Lognormal Model



Lognormal Model



Properties of Expected Value

Call Option Payoff:

- Option to Buy asset at time T at
  Strike Price: K
- X: price of asset at time T
- Payoff: $C = (X - K)^{+} = max(0, X - K)$

Theorem: Let $f(x)$ be the probability density of X, $F(x) = \int_{-\infty}^{x} f(t)dt$ be the cumulative distribution function. If K is a constant and X has finite variance, then

$$E[(X-K)^{+}] = \int_K^{\infty} \left(\int_x^{\infty} f(t)dt\right) dx = \int_K^{\infty}[1 - F(x)]dx.$$

Proof:

$$\begin{aligned} E[(X-K)^{+}] &= \int_K^{\infty}(x-K)f(x)dx \\ &= \lim_{M\to\infty} \int_K^{M}(x-K)f(x)dx \end{aligned}$$

Integrate by parts:

$$\begin{aligned} u &= x - K & \quad v &= -[1 - F(x)] \\ du &= dx & \quad dv &= f(x)dx \end{aligned}$$

Properties of Expected Value

Corollary 1: If X is a normal random variable with mean μ and standard deviation σ, and K is a constant, then

$$E[(X-K)^+] = (\mu - K)\Phi(\tfrac{\mu-K}{\sigma}) + \tfrac{\sigma}{\sqrt{2\pi}} e^{\frac{(K-\mu)^2}{2\sigma^2}}$$

where $\Phi(x)$ is the cumulative distribution function of the standard normal distribution (mean 0, standard deviation 1).

Corollary 2: If X is a log-normal(μ, σ) random variable and K is a constant, then

$$E[(X-K)^+] = e^{\mu+\sigma^2/2}\Phi(\tfrac{\mu-\ln K}{\sigma} + \sigma) - K\Phi(\tfrac{\mu-\ln K}{\sigma})$$

Moment-Generating Function: Definition/Theory

Definition: The **moment-generating function** of a random variable X : $M_X(t) = E[e^{tX}] = \sum_{k=0}^{\infty} \frac{t^k}{k!} m_k$

Theorem: Let X be a random variable with moment-generating function $M_X(t)$ and cumulative distribution function $F_X(x)$.

1. If Y is a a random variable satisfying $M_Y(t) = M_X(t)$ for all t, then X and Y have identical distributions, i.e., the cumulative distribution functions are the same.
2. Let $X_1, X_2, \ldots, X_n$ be a sequence of random variables such that
$$\lim_{i \to \infty} M_{X_i}(t) = M_X(t)$$
for all t. Then X_i converges to X in distribution, i.e., $\lim_{i \to \infty} F_{X_i}(x) = F_X(x)$, for all real x.

Moment Generating Functions (MGFs)

Normal Distribution: For $X \sim N(\mu, \sigma^2)$

$$M_X(t) = E[e^{tX}] = e^{\mu t + \sigma^2 t^2/2}$$

- Special case: standard normal $Z \sim N(\mu = 0, \sigma^2 = 1)$

$$M_X(t) = E[e^{tZ}] = e^{\frac{1}{2}t^2}$$

MGF of a Linear Transformation

- Random variable X has mgf $M_X(t)$.
- Define $Y = \mu + (\sigma \times X)$, for constants μ, σ
- The MGF of Y is

$$\begin{aligned} M_Y(t) &= E[e^{tY}] \\ &= E[e^{t(\mu + \sigma X)}] \\ &= e^{t\mu} E[e^{t\sigma X}] \\ &= e^{t\mu} M_X(t\sigma) = e^{\mu t + \frac{\sigma^2}{2} t^2} \end{aligned}$$

(Note: compute moments of Lognormal Distribution using Normal MGF)

Exercise: Find the skewness of a *lognormal*(μ, σ^2) random variable.

Exercise: Show that Kurtosis $\kappa = +3$ for a *Normal*(μ, σ^2) distribution.

Definition: A random variable X is **leptokurtic** if $\kappa > 3$.

Linear Transformations:

Consider a random variable X with mean μ and variance σ^2.
The linear transformation of X: $Y = a + bX$ (constants a and b)

- $E[Y] = a + bE[X] = a + b\mu$
- $Var[Y] = b^2 Var[X] = b^2\sigma^2$.

Exercise: If the skewness of X is γ, what is the skewness of Y?
Exercise: If the kurtosis of X is κ, what is the kurtosis of Y?

Probability Concepts for Several Random Variables

Independent Random Variables / Events

- Two random variables

  X (with sample space $\mathcal{X}$ and pmf/density $f_X(x)$)

  Y (with sample space $\mathcal{Y}$ and pmf/density $f_Y(y)$)

- X and Y are **independent** if

  $$P(\{X \in A\} \cap \{Y \in B\}) = P(\{X \in A\}) \times P(\{Y \in B\})$$

  for all $A \subset \mathcal{X}$ and all $B \subset \mathcal{Y}$

- If X and Y are independent, then the density/pmf function of the joint distribution of (X, Y) is

  $$f_{X,Y}(x, y) = f_X(x) f_Y(y)$$

Covariance and Correlation

Definitions

- The **covariance** of two random variables X and Y is
$$cov(X, Y) = E[(X - E[X])(Y - E[Y])]$$
Note: $cov(X, X) = var(X)$.

- The **correlation** of two random variables X and Y is
$$cor(X, Y) = \frac{cov(X, Y)}{\sqrt{var(X)var(Y)}}$$

Note: If X and Y are independent, then
$$cov(X, Y) = 0 \text{ and } cor(X, Y) = 0..$$
(If $cor(X, Y) = 0$, then X and Y may not be independent!)

Random Vectors and Covariance Matrices

- Random variables: $X_1, X_2, \ldots, X_n$ with
$$\mu_j = E[X_j],\ j = 1, \ldots, n$$
$$\sigma_{i,j} = cov(X_i, X_j),\ i, j = 1, \ldots, n$$
- Random vector, mean vector
$$\vec{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix},\ E[\vec{X}] = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} = \vec{\mu},$$
- Covariance matrix $\Sigma = ||\sigma_{i,j}||$ with $\sigma_{i,j} = cov(X_i, X_j)$
$$\underset{(n \times n)}{\Sigma} = cov(\vec{X}) = E[(\vec{X} - E[\vec{X}])(\vec{X} - E[\vec{X}])^T]$$
- For $\vec{a} = [a_1, a_2, \ldots, a_n]^T$ (constant vector) define
$$Y = \vec{a}^T\vec{X} = a_1X_1 + a_2X_2 + \cdots + a_nX_n$$

Random Vectors and Covariance Matrices

(continued)

- $E[Y] = E[\vec{a}^T\vec{X}] = \vec{a}^T E[\vec{X}] = \vec{a}^T\vec{\mu}$
- $$\begin{aligned} var(Y) &= E[(Y - E[Y])^2] \\ &= E[(\vec{a}^T X - \vec{a}^T\vec{\mu})^2] \\ &= E[(\vec{a}^T(\vec{X} - \vec{\mu}))^2] \\ &= E[\vec{a}^T[(\vec{X} - \vec{\mu})][(\vec{X} - \vec{\mu})]^T\vec{a}] \\ &= \vec{a}^T E[[(\vec{X} - \vec{\mu})][(\vec{X} - \vec{\mu})]^T]\vec{a} \\ &= \vec{a}^T[cov(\vec{X})]\vec{a} \\ &= \vec{a}^T\Sigma\vec{a} \\ &= \textstyle\sum_i \sum_j a_i a_j \sigma_{i,j} \\ &= \sum_i a_i^2 var(X_i) + 2\sum_{i<j} a_i a_j cov(X_i, X_j) \end{aligned}$$

Principal Components Analysis (PCA)

- An $m-$variate random variable:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}, \text{ with } E[\mathbf{x}] = \boldsymbol{\alpha} \in \Re^m, \text{ and } Cov[\mathbf{x}] = \boldsymbol{\Sigma}_{\ (m \times m)}$$

- Eigenvalues/eigenvectors of $\boldsymbol{\Sigma}$:
  - $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq 0$: m eigenvalues.
  - $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_m$: m orthonormal eigenvectors:

$$\begin{aligned} \boldsymbol{\Sigma}\boldsymbol{\gamma}_i &= \lambda_i \boldsymbol{\gamma}_i, \quad i = 1, \ldots, m \\ \boldsymbol{\gamma}_i' \boldsymbol{\gamma}_i &= 1, \quad \forall i \\ \boldsymbol{\gamma}_i' \boldsymbol{\gamma}_{i'} &= 0, \quad \forall i \neq i' \end{aligned}$$

  - $\boldsymbol{\Sigma} = \sum_{i=1}^{m} \lambda_i \boldsymbol{\gamma}_i \boldsymbol{\gamma}_i'$
- Principal Component Variables:

$$p_i = \boldsymbol{\gamma}_i'(\mathbf{x} - \boldsymbol{\alpha}), \quad i = 1, \ldots, m$$

Principal Components in Vector/Matrix Form

- $m-$Variate $\mathbf{x}$: $\quad E[\mathbf{x}] = \boldsymbol{\alpha},\ Cov[\mathbf{x}] = \boldsymbol{\Sigma}$
- $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}'$, where
$$\begin{array}{rcl} \boldsymbol{\Lambda} & = & diag(\lambda_1, \lambda_2, \ldots, \lambda_m) \\ \boldsymbol{\Gamma} & = & [\gamma_1 : \gamma_2 : \ \cdots \ : \gamma_m] \\ \boldsymbol{\Gamma}'\boldsymbol{\Gamma} & = & \mathbf{I}_m \end{array}$$
- $\mathbf{p} = \begin{bmatrix} p_1 \\ \vdots \\ p_m \end{bmatrix} = \boldsymbol{\Gamma}'(\mathbf{x} - \boldsymbol{\alpha})$, $m-$Variate PC variables
$$\begin{array}{rcl} E[\mathbf{p}] & = & E[\boldsymbol{\Gamma}'(\mathbf{x} - \boldsymbol{\alpha})] = \boldsymbol{\Gamma}' E[(\mathbf{x} - E[\mathbf{x}])] = \mathbf{0}_m \\ Cov[\mathbf{p}] & = & Cov[\boldsymbol{\Gamma}'(\mathbf{x} - \boldsymbol{\alpha})] = \boldsymbol{\Gamma}' Cov[\mathbf{x}]\boldsymbol{\Gamma} \\ & = & \boldsymbol{\Gamma}'\boldsymbol{\Sigma}\boldsymbol{\Gamma} = \boldsymbol{\Gamma}'(\boldsymbol{\Gamma}\boldsymbol{\lambda}\boldsymbol{\Gamma}')\boldsymbol{\Gamma} = \boldsymbol{\Lambda} \end{array}$$
- $\mathbf{p}$ is a vector of zero-mean, uncorrelated random variables that provides an *orthogonal basis* for $\mathbf{x}$.

Principal Components Analysis

$m-$Variate x in Principal Components Form

- $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} = \boldsymbol{\alpha} + \boldsymbol{\Gamma}\mathbf{p}$, where $E[\mathbf{p}] = \mathbf{0}_m$, $Cov[\mathbf{p}] = \boldsymbol{\Lambda}$
- Partition $\boldsymbol{\Gamma} = [\boldsymbol{\Gamma}_1 \boldsymbol{\Gamma}_2]$ where $\boldsymbol{\Gamma}_1$ corresponds to the K ($< m$) largest eigenvalues of $\boldsymbol{\Sigma}$.
- Partition $\mathbf{p} = \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \end{bmatrix}$ where $\mathbf{p}_1$ contains the first K elements.
- $\mathbf{x} = \boldsymbol{\alpha} + \boldsymbol{\Gamma}_1\mathbf{p}_1 + \boldsymbol{\Gamma}_2\mathbf{p}_2 = \boldsymbol{\alpha} + B\mathbf{f} + \boldsymbol{\epsilon}$

  where

$$\begin{array}{rcll} B & = & \boldsymbol{\Gamma}_1 & {\scriptstyle (m \times K)} \\ \mathbf{f} & = & \mathbf{p}_1 & {\scriptstyle (K \times 1)} \\ \boldsymbol{\epsilon} & = & \boldsymbol{\Gamma}_2\mathbf{p}_2 & {\scriptstyle (m \times 1)} \end{array}$$

  Like factor model except $Cov[\boldsymbol{\epsilon}] = \boldsymbol{\Gamma}_2\boldsymbol{\Lambda}_2\boldsymbol{\Gamma}_2'$, where $\boldsymbol{\Lambda}_2$ is diagonal matrix of last $(m - K)$ eigenvalues.

Empirical Principal Components Analysis

The principal components analysis of

$$\mathbf{X} = [\mathbf{x}_1 : \cdots \mathbf{x}_T]_{(m \times T)}$$

consists of the following computational steps:

- Component/row means : $\bar{\mathbf{x}} = (\frac{1}{T})\mathbf{X}\mathbf{1}_T$
- 'De-meaned' matrix: $\mathbf{X}^* = \mathbf{X} - \bar{\mathbf{x}}\mathbf{1}_T'$
- Sample covariance matrix: $\hat{\mathbf{\Sigma}}_x = \frac{1}{T}\mathbf{X}^*(\mathbf{X}^*)'$
- Eigenvalue/vector decomposition: $\hat{\mathbf{\Sigma}}_x = \hat{\mathbf{\Gamma}}\hat{\mathbf{\Lambda}}\hat{\mathbf{\Gamma}}'$
  yielding estimates of $\mathbf{\Gamma}$ and $\mathbf{\Lambda}$.
- Sample Principal Components:

$$\mathbf{P} = [\mathbf{p}_1 : \cdots : \mathbf{p}_T] = \hat{\mathbf{\Gamma}}'\mathbf{X}^*._{(m \times T)}$$

Empirical Principal Components Analysis

PCA Using Singular Value Decomposition

Consider the Singular Value Decomposition (SVD) of the de-meaned matrix:

$$\mathbf{X}^* = \mathbf{VDU}'$$

where

- $\mathbf{V}$: $(m \times m)$ orthogonal matrix, $\mathbf{VV}' = \mathbf{I}_m$.
- $\mathbf{U}$: $(m \times T)$ row-orthonormal matrix, $\mathbf{UV}' = \mathbf{I}_m$.
- $\mathbf{D}$: $(m \times m)$ diagonal matrix, $\mathbf{D} = diag(d_1, \dots, d_m)$ with $d_1 \geq d_2 \geq \cdots \geq 0$.

Exercise: Show that

- $\hat{\mathbf{\Lambda}} = \frac{1}{T}\mathbf{D}^2$
- $\hat{\mathbf{\Gamma}} = \mathbf{V}$
- $\mathbf{P} = \hat{\mathbf{\Gamma}}'\mathbf{X}^* = \mathbf{DU}'$

Alternate Definition of PC Variables

Given the $m-$variate $\mathbf{x} : E[\mathbf{x}] = \alpha$ and $Cov[\mathbf{x}] = \mathbf{\Sigma}$

- Define the **First Principal Component Variable** as
  $$p_1 = \mathbf{w}'\mathbf{x} = (w_1 x_1 + w_2 x_2 + \cdots + w_m x_m)$$
  where the coefficients $\mathbf{w}= (w_1, w_2, \ldots, w_m)'$ are chosen to
  maximize: $Var(p_1) = \mathbf{w}'\mathbf{\Sigma}_x\mathbf{w}$
  subject to: $|\mathbf{w}|^2 = \sum_{i=1}^m w_i^2 = 1.$
- Define the **Second Principal Component Variable** as
  $$p_2 = \mathbf{v}'\mathbf{x} = (v_1 x_1 + v_2 x_2 + \cdots + v_m x_m)$$
  where the coefficients $\mathbf{v}= (v_1, v_2, \ldots, v_m)'$ are chosen to
  maximize: $Var(p_2) = \mathbf{v}'\mathbf{\Sigma}_x\mathbf{v}$
  subject to: $|\mathbf{v}|^2 = \sum_{i=1}^m v_i^2 = 1$, **and** $\mathbf{v}'\mathbf{w} = 0$.
- Etc., defining up to p_m, The coefficient vectors are given by
  $$[\mathbf{w} : \mathbf{v} : \cdots] = [\gamma_1 : \gamma_2 : \cdots] = \mathbf{\Gamma}$$

Principal Components Analysis

Further Details

- PCA provides a decomposition of the **Total Variance**:

$$
\begin{aligned}
\text{Total Variance }(\mathbf{x}) &= \sum_{i=1}^{m} Var(\mathbf{x}_i) = trace(\boldsymbol{\Sigma}_x) \\
&= trace(\boldsymbol{\Gamma\Lambda\Gamma}') = trace(\boldsymbol{\Lambda\Gamma}'\boldsymbol{\Gamma}) = trace(\boldsymbol{\Lambda}) \\
&= \sum_{k=1}^{m} \lambda_k \\
&= \sum_{k=1}^{m} Var(p_k) \\
&= \text{Total Variance }(\mathbf{p})
\end{aligned}
$$

- The transformation from $\mathbf{x}$ to $\mathbf{p}$ is a change in coordinate system which shifts the origin to the mean/expectation $E[\mathbf{x}] = \boldsymbol{\alpha}$ and rotates the coordinate axes to align with the Principal Component Variables. Distance in the space is preserved (due to orthogonality of the rotation).

Chi-Square Distributions

Definition. If $Z \sim N(0,1)$ (Standard Normal r.v.) then

$$U = Z^2 \sim \chi_1^2,$$

has a **Chi-Squared distribution with 1 degree of freedom.**

Properties:

- The density function of U is:
$$f_U(u) \;=\; \frac{u^{-1/2}}{\sqrt{2\pi}} e^{-u/2}, \;\; 0 < u < \infty$$
- Recall the density of a $Gamma(\alpha, \lambda)$ distribution:
$$g(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \; x > 0,$$
So U is $Gamma(\alpha, \lambda)$ with $\alpha = 1/2$ and $\lambda = 1/2$.
- Moment generating function
$$M_U(t) = E[e^{tU}] = [1 - t/\lambda]^{-\alpha} = (1 - 2t)^{-1/2}$$

Chi-Square Distributions

Definition. If $Z_1, Z_2, \ldots, Z_n$ are i.i.d. $N(0,1)$ random variables
$$V = Z_1^2 + Z_2^2 + \ldots Z_n^2 \text{ has a } \chi_n^2 \text{ distribution}$$

Properties

- A Chi-Square r.v. V (n degrees of freedom) equals
  $$V = U_1 + U_2 + \cdots + U_n$$
  where $U_1, \ldots, U_n$ are i.i.d χ_1^2 r.v.
- Moment generating function
  $$\begin{aligned} M_V(t) &= E[e^{tV}] = E[e^{t(U_1+U_2+\cdots+U_n)}] \\ &= E[e^{tU_1}] \cdots E[e^{tU_n}] = (1-2t)^{-n/2} \end{aligned}$$
- Because U_i are i.i.d. $Gamma(\alpha = 1/2, \lambda = 1/2)$ r.v.,s
  $$V \sim Gamma(\alpha = n/2, \lambda = 1/2).$$
- Density function: $f(v) = \dfrac{1}{2^{n/2}\Gamma(n/2)} v^{(n/2)-1} e^{-v/2}, v > 0.$

  (α is the **shape parameter** and λ is the **scale parameter**)

Student's t Distribution and Fisher's F Distribution

Definition. For *independent* r.v.'s Z and U where

- $Z \sim N(0,1)$
- $U \sim \chi^2_r$

the distribution of $T = Z/\sqrt{U/r}$ is the

t **distribution with** r **degrees of freedom**.

Definition. For *independent* r.v.'s U and V where

- $U \sim \chi^2_m$ and $V \sim \chi^2_n$

the distribution of $F = \dfrac{U/m}{V/n}$ is the

F **distribution with** m **and** n **degrees of freedom**.

(notation $F \sim F_{m,n}$)

Properties

- $E[F] = E[U/m] \times E[n/V] = 1 \times n \times \frac{1}{n-2} = \frac{n}{n-2}$ (for $n > 2$).
- If $T \sim t_r$, then $T^2 \sim F_{1,r}$.

Law of Large Numbers

Theorem (Weak Law of Large Numbers - WLLN).
Suppose $X_1, X_2, \ldots, X_n, \ldots$ are i.i.d.
(independent and identically distributed)
with $E[X_i] = \mu$ and $var(X_i) = \sigma^2$.
Define $\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$.
Then $\bar{X}_n \xrightarrow{pr} \mu$, i.e., for any $\epsilon > 0$

$$\lim_{n\to\infty} P(|\bar{X} - \mu| > \epsilon) = 0.$$

Proof: Apply Chebycheff's Inequality

- $var(\bar{X}_n) = \frac{\sigma^2}{n}$
- $\epsilon^2 P(|\bar{X}_n - \mu| > \epsilon) \leq var(\bar{X}_n)$
- $\implies P(|\bar{X}_n - \mu| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2}\frac{1}{n} \to 0$

Central Limit Theorem

Theorem (Central Limit Theorem). Let $X_1, X_2, \ldots$ be i.i.d. random variables with

$$E[X_i] = 0 \text{ and } var[X_i] = \sigma^2, \text{ and MGF } M(t) = E[e^{tX_i}]$$

Then the sequence of random variables $Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$ converges in distribution to the normal distribution $N(0, \sigma^2)$.

Proof: Evaluate the MGF of Z_n:

$$\begin{aligned} M_{Z_n}(t) = E[e^{tZ_n}] &= E[e^{t\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i}] \\ &= \prod_{i=1}^n E[e^{\frac{t}{\sqrt{n}} X_i}] \\ &= \prod_{i=1}^n M(\tfrac{t}{\sqrt{n}}) = [M(\tfrac{t}{\sqrt{n}})]^n \end{aligned}$$

Apply Taylor Series to the MGF of X:

$$\begin{aligned} M(\tfrac{t}{\sqrt{n}}) &= 1 + E[X](\tfrac{t}{\sqrt{n}}) + \tfrac{E[X^2]}{2}(\tfrac{t}{\sqrt{n}})^2 + O((\tfrac{t}{\sqrt{n}})^3) \\ &= 1 + \tfrac{\sigma^2 t^2/2}{n} + o(\tfrac{t^2}{n}) \\ \Longrightarrow \quad M_{Z_n}(t) &= [M(\tfrac{t}{\sqrt{n}})]^n \to e^{\sigma^2 t^2/2} \text{ the MGF of } N(0, \sigma^2) \end{aligned}$$