Boston University
Questrom School of Business

MF793 – Fall 2021

Eric Jacquier

Random Variables and Distributions

1 Random variable (RV)

- Definition,
- Discrete RV
- Cumulative distribution function (CDF)
- Probability density function (PDF)
    - Kernel and normalization constant of a pdf
- Multivariate RVs
    - Bayes Theorem for RVs, i.id samples

2 Location estimates of a RV

- Mean Squared Error criterion (MSE), loss function
- Mean, Median, Mode

3 Higher moments of a distribution

- Variance, covariance, correlation
- Basic moment relations
    - Variance of a portfolio
    - Chebishev
- Skewness, kurtosis, central vs non-central moments
- Moment generating function

4 Function of a RV

- Finding the CDF and PDF of a function of a RV
    - Monotone and non-monotone transformations
- Application: the Probability Inverse transform

1. Random variables (RV)

1.1 Definition

We often simplify the description of complex events by the use of a random variable.

This random variable (RV) summarizes what we are interested in

Definition: **Random Variable**:
Function from the sample space to the real numbers (can be multivariate)

- Sum of two dice: $x = d_1+d_2$,
- Multivariate RV: keep track of each die: (d_1,d_2)
- Number of defective bulbs produced in a batch
- Fraction of defective bulbs produced
- \$/£, \$/¥ exchange rate
- IBM stock price ... or stock return, multivariate: (R_{IBM},R_{AAPL})
- Stock Index, CPI, GDP level ... or their % change
- Rare events: jump, crash a discrete 0/1 integer RV

Question:
Why do we always model the time series behavior of stock returns (**not** prices)?
exchange rate % change (**not** levels), inflation (**not** CPI),
growth rate of GDP (**not** level) ?

1.2 Discrete Distributions

- X can take discrete values: x_1, x_2, …., x_n
  … with probabilities: p_1, p_2, …., p_n

  $\{p_i\}$ is a ***probability function***: $\sum_{i=1}^{n} p_i = 1, \quad p_i \geq 0 \;\; \forall i, \quad p(X = x_i \cup X = x_j) = p_i + p_j$

  also denoted probability ***mass*** function

- … are useful for
  - Insurance
  - Some Economic scenario approaches
  - Capital budgeting (often uses scenarios)
  - Dividend forecasts (firm will / will not increase dividends)

- Discrete distributions **are not practical** for economic and financial series levels, prices, growth rates and returns

  - There a just too many possible outcomes

  - **Continuous distributions** are more practical in these cases

1.3 Cumulative Distribution Function (CDF) of a RV

Definition: **Cumulative Distribution Function** (CDF) of a RV X

Denoted by $F_X(X)$, it is defined by

$$\boldsymbol{F_X(x_0) = P_X(X \leq x_0), \qquad \forall x_0 \in D_x}$$

X: the random variable
x_0: a specific value in D_x the domain of X

Observation: We defined the cumulative function first…
because it is well defined for both continuous and discrete RVs

Properties:
Denote LB and UB, the lower and upper bounds of the domain D_x.
CDF has 3 obvious properties:

$$\lim_{x0 \to UB} F_X(x_0) = 1 \qquad \lim_{x0 \to LB} F_X(x_0) = 0 \qquad F_X(x_0) \text{ non-decreasing in } x_0$$

Observation: CDF does not need to be continuous, it can have jumps
One more condition not written above is that it is right-continuous

Definition: A RV is **continuous** if its CDF is continuous
A RV is **discrete** if its CDF has jumps

1.4 Probability mass and probability density functions

- Discrete Distribution has a **Probability Mass Function** *$p(x_0)$ = Prob.(X=x_0)*

  CDF is the cumulative sum of Probability Mass up to <u>and including</u> x_0.

  We often first describe the Probability Mass, then the CDF if we need it.

- Continuous Distribution: **Probability Density Function (PDF)** $\boldsymbol{f(x_0) = \frac{d}{dx_0} F(x_0)}$

  Or … the CDF is the partial integral of the PDF up to x_0

$$\boldsymbol{F(x_0) = \int_{LB}^{x_0} f(t)dt}$$

We **can** write the analytical formula for the **PDF** of known densities.

We most often **cannot** write an analytical formula for their **CDF**

Interval probability: $\boldsymbol{P(a < X < b)} = \boldsymbol{F(b)} - \boldsymbol{F(a)} = \int_a^b f(t)dt$

α% Confidence (or credibility) interval $\boldsymbol{P(a < X < b) = \alpha}$

- To be a proper pdf, a pdf function must:

    i. be ≥0 everywhere to guarantee that all its partial integrals, i.e. interval probabilities, are ≥0

    ii. integrate to 1 over the domain: $\int_{LB}^{UB} f(t)dt = 1$

- Say you have a function *f(t)* with $\int_{LB}^{UB} f(t)dt = \frac{1}{C} \neq 1$?

  Then: $\int_{-\infty}^{+\infty} C\, f(t)dt = 1$ [1]

  It works! **A non-negative function with finite integral can be rescaled as a proper pdf.**

- **Normalization Constant and Kernel of a PDF**

  If we have *f(x)*, we can find ***C*** with the integration above. Then a proper pdf $p(x_0)$ is:

p(x)	**=**	***C***	**x**	***f(x)***
pdf of x		**Normalization Constant**		**Kernel**

  f(x): **Kernel**. As the sole function of x it completely determines the shape of the density

  C: Computed by solving [1]. C scales the kernel to sum to 1, **C is not a function of x**

1.5 Multivariate RV: joint density, conditional density, marginal density

- CDF of a **multivariate** random variable (X,Y), aka the **joint** CDF

$$F_{X,Y}(x_0, y_0) = P_{XY}(X \le x_0, Y \le y_0), \qquad \forall x_0, y_0 \in D_{xy}$$

- PDF of a multivariate density: $p_{X,Y}(x_0,y_0) = \frac{d}{dx_0\, dy_0} F(x_0, y_0)$

$$F_{X,Y}(x_0, y_0) = \int_{LBx}^{x_0} \int_{LBy}^{y_0} p_{X,Y}(t,s)\,dt\,ds$$

- **Bayes theorem** applies to pdfs the same way it applies to probabilities of outcomes:

$$p(x,y) = p(y|x)\,p(x) = p(x|y)\,p(y)$$

- Same for the Total Probability Theorem: $P(A) = \sum_{i=1}^{n} P(A \text{ and } B_i) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$

  p(x) the **marginal density** of x is: $p(x) = \int p(x,y)\,dy \quad = \int p(x|y)\; p(y)\; dy$

- **I**ndependently **I**dentically **D**istributed RVs: **i. i. d. random samples**

  If n RVs (X_1, ... , X_i, ... ,X_n) are independently distributed with the same pdf $p(X_i)$, then:

$$p(X_1, \ldots, X_n) = \prod_{i=1}^{n} p_i(X_i) = \prod_{i=1}^{n} p(X_i)$$

2 Location Estimates of a Distribution

- For a RV, X with pdf p(X), what number would we use to describe its location?

- **This has nothing to do with estimating parameters from data !** 😡
  Here we define potentially interesting parameters as ways to describe a distribution.

- We do **not** say (yet) how to best use data to estimate these parameters (will be next notes).

2.1 The MEAN - fundamental property (why we like it !)

Want a number θ which is "best" at predicting where X is likely to be.

What is "best"? Need a criterion for "best".

Definition: **Mean Squared Error** (MSE) of prediction of RV X: **E[$(\tilde{X} - \theta)^2$]**

Goal: Find θ with the lowest MSE, minimize E[$(\tilde{X} - \theta)^2$]: E[$(\tilde{X} - \theta)^2$] = $\int (x-\theta)^2 p(x)dx$

We can do it two ways

- Set first derivative with respect to θ equal to 0.

$$0 = \int -2(x-\theta)p(x)dx \quad =$$

$$\Rightarrow \quad \boldsymbol{\theta = E(X)}$$

Note: Bounds are not a function of θ, otherwise we would need to use full Leibniz rule

- Or: Let $\mu = E(X)$ $\quad E[(X - \theta)^2] = E[(X - \mu + \mu - \theta)^2]$

$$= E[(X - \mu)^2] + E[(\mu - \theta)^2] + E[(X-\mu)(\mu - \theta)]$$

$$=$$

The Mean minimizes the mean squared error of prediction

- Note again: **Not** about estimating the mean from data, this is a property of the true mean as it describes where the RV may be. 😡

2.2 Other location estimates: The Median

Median P(X ≤ m) = P(X ≥ m) = 0.5 (Continuous distribution)

- Convenient property:

  If m is the median of x, then g(m) is the median of y = g(x) for g monotone

  Median is invariant to transformations

- What is the median best for ?

  The Median minimizes the Mean Absolute Error (MAE) of prediction E|X - θ|

MAE = $\int_{LB}^{UB} |x-\theta|\, p(x)\, dx$ = $\int_{LB} \quad (x-\theta) \quad p(x)\, dx + \int^{UB} \quad (x-\theta) \quad p(x)\, dx$

$\frac{\partial MAE}{\partial \theta} =$

Bounds are now a function of t, We need to use Leibniz rule:

$$\frac{d}{d\theta}\int_{L(\theta)}^{U(\theta)} f(x,\theta)dx \ = \int_{L(\theta)}^{U(\theta)} f'(x,\theta)dx + f(x = U(\theta),\theta)\ U'(\theta) - f(x = L(\theta),\theta)\ L'(\theta)$$

$$\frac{\partial MAE}{\partial \theta} =$$

2.3 Third location estimates: **Mode:** **p(Mode) = Max$_x$ p(x)**

Can show: The mode minimizes the 0/1 prediction error function: **L(X,θ) = $I_{x\neq\theta}$**

I: indicator function (I=1 if X= θ, 0 otherwise)

L(X, θ) : Loss function of using θ as a predictor of X

CONCLUSION:

Three Loss functions => three different optimal location parameters.

The loss function is a fundamental tool in Decision Theory

3 Spread of a Distribution: Second Moments

3.1 Variance, covariance, correlation

- Variance: $V(x) = E(x-\mu)^2 = \int (x-\mu)^2\, p(x)\, dx = \sigma_x^2$

  $V(x) = E(x^2) - [E(x)]^2$ prove it

- Covariance: $Cov(X,Y) = E(X - \mu_X)(Y - \mu_Y)$ Expectation over joint density $p_{XY}(x,y)$

  Notation: we write $\sigma_{X,Y}$

  $\sigma_{X,Y} = E(XY) - \mu_X \mu_Y$ prove it

  $\sigma_x^2 = Cov(x,x)$

Covariance has an intuitive sign, **not an intuitive magnitude**

- Correlation: $Cor(X,Y) = \sigma_{X,Y} / \sigma_X \sigma_Y$ Notation: we write ρ_{XY}

Schwartz inequality: $[E(UV)]^2 \leq E(U^2)E(V^2)$ No proof

Then we can show: $[Cov(X,Y)]^2 \leq Var(X)\, Var(Y)$ prove it

$|\rho_{XY}| < 1$!

Correlation has an intuitive magnitude on [0,1]

3.2 Basic moment relations

- Mean is a linear operator **$E(a\tilde{X} + b\tilde{Y} + c) = a\,E(\tilde{X}) + b\,E(\tilde{Y}) + c$**

- Variance is a quadratic operator **$Var(a\tilde{X} + b) = a^2\,Var(\tilde{X})$**

- Standard deviation **$\sigma_X = \sqrt{Var(X)}$** is linear in X.

- Standardization: **if $R \sim (\mu, \sigma)$ then $(R-\mu)/\sigma \sim (0,1)$**

- Covariance is a bi-linear operator $Cov(a\tilde{X}, b\tilde{Y}) = ab\,\sigma_{X,Y}$

- Correlation is a measure of **linear** dependence: **if $Y = -|a|\,X + b$, then $\rho_{XY} = -1$**

  Prove it: compute ρ_{XY}

- (X,Y) independent => $\sigma_{X,Y} = \rho_{XY} = 0$

  Converse is not true ! Think of $Y = X^2$

- Variance of a combination:

$$Var(a\,X + b\,Y\,) = a^2\,\sigma_x^2 + b^2\,\sigma_x^2 + 2\,ab\,\sigma_{xy}$$

$$E\left[\left(aX + bY - a\mu_x - b\mu_y\right)^2\right] = E\left[\left(aX - a\mu_x\right)^2 + \left(bY - b\mu_y\right)^2 + 2\left(aX - a\mu_x\right)\left(bY - b\mu_y\right)\right]$$

Variance of a portfolio's return:

$$R_P = \sum_{1}^{n} w_i\,R_i$$

$$Var\left(\sum_{i=1}^{n} w_i R_i\right) = \sum_{1}^{n} w_i^2 Var(R_i) + \sum_{i}\sum_{j\neq i} w_i w_j Cov(R_i\,, R_j)$$

Matrix form **$Var\,(w^T R) = w^T\,\Omega\,w$**

Ω: covariance matrix of the vector of random variables $R = (R_1, \ldots, R_n)$,

$$\Omega = E[\,(\,R - \mu_R)\,(R - \mu_R)^T\,]$$

- **Moments vs Central Moments**

  - Non central moment: $E[X^k]$
  - Central moment about the mean: $\mathbf{E[(X-\mu)^k]}$

3.3 Less basic (but no less important) relations

$$E(X) = \int_0^\infty x\,\varphi(x)\,dx = \int_0^c + \int_c^\infty \geq \int_c^\infty c\,\varphi(x)\,dx \to c\underbrace{\int_c^\infty \varphi(x)\,dx}_{P(X>c)}$$

- **Inequalities** **$X \sim (\mu, \sigma) > 0$**
  - Markov inequality $P(X > c) < \mu / c$, for $c > 0$ [1] **Prove it**

    $P(X \geq k\mu) < \frac{1}{k}$
  - **Chebyshev** **$P(|X-\mu| \geq c\sigma) \leq 1/c^2$** **Prove from** [1]

    Often not very tight: Try c = 1, 2, 3 and compare to normal density.

    .. But it is important for convergence results

    .. And it applies no matter the density of X.

  - **Jensen** if g(x) is a **convex** function, **$E(g(X)) \geq g(E(X))$**

$$E(X^2) \geq E(X)^2\ !$$
$$E(U(w)) \leq U(E(w))$$

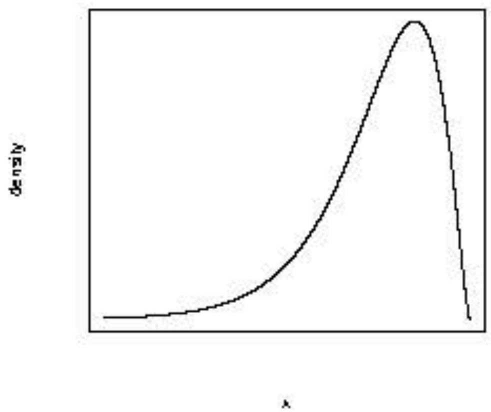
- **Iterated Expectation and other rules**
  - **$E(Y) = E_X[\,E_{Y|X}(Y \mid X)\,]$**
  - $Var(Y) = E_X[\,Var(Y|X)\,] + Var_X[E(Y|X)]$
  - If $(\ldots X_i \ldots)$ are independent $E(\prod_{i=1}^n X_i) = \prod_1^n E(X_i)$ prove it

偏度与峰度

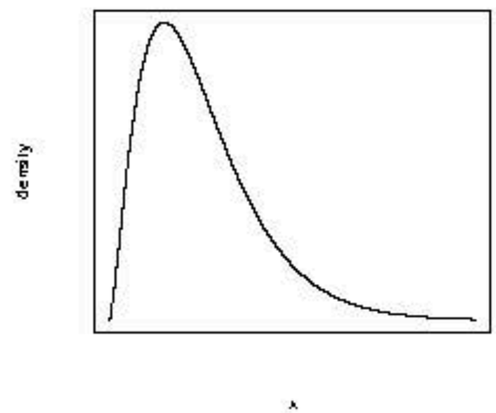
3.4 Higher Moments: Skewness and Kurtosis

- **Skewness:** $Sk = E\left(\frac{R-\mu}{\sigma}\right)^3$ No asymmetry: $Sk = 0$

Negative Skewness $Sk < 0$



Positive Skewness $Sk > 0$



- We like positive skewness
  We do not like negative skewness: Higher volatility on negative outcomes.
  Leverage Effect

- **Kurtosis** $K = E\left(\frac{R-\mu}{\sigma}\right)^4$ Normal Distribution: $K = 3$

- **Fat tails: K > 3.** more risky than the normal, important for risk management especially at short horizons (daily, weekly, ..)

3.5 Moment Generating Function

- $M_X(t) = E[e^{tX}]$ is the ***moment generating function*** of X

- Result: $$E[X^n] = \frac{d^n}{dt^n} M_X(t)\Big|_{t=0}$$

  Proof: $\frac{d}{dt} M_X(t) = \frac{d}{dt}\int_L^U e^{tx} f(x)dx = \int_L^U \frac{d}{dt}(e^{tx}) f(x)dx = \int_L^U x e^{tx} f(x) dx$ [1]

  $= E[Xe^{tX}]\Big|_t$

  At t=0, we have: $\frac{d}{dt} M_X(0) = E[Xe^0] = E(X)$

- Recall from integration calculus:

  1. Exchanging integral and differentiation (aka differentiating inside the integral) requires smoothness of the derivative. See Casella & Berger

  2. Bounds in [1] are not a function of t, special case of Leibniz rule. Don't forget the general Leibniz rule:
  $$\frac{d}{d\theta}\int_{L(\theta)}^{U(\theta)} f(x,\theta)dx = \int_{L(\theta)}^{U(\theta)} f'(x,\theta)dx + f(x = U(\theta),\theta)\, U'(\theta) - f(x = L(\theta),\theta)\, L'(\theta)$$

- The MGF is often a very convenient way to compute moments of densities or functions of densities.
  But .. a density may have moments but no well defined MGF !! (e.g., lognormal density)

4 Distribution of a Function of a RV

4.1 Principle of change of variable

- Often need the density of a function g of a RV: Y = g(X)

  We estimate a parameter, we want a function of the parameter
  We want to forecast a function of a RV for which we have the predictive density

- How to do this: Associate an inverse mapping to g(X), from Y back to X

  $$g^{-1}(\{y\}) = \{x: g(x) = y\}$$

  and

  $$P(Y \in A) = P(g(X) \in A) = P(X \in g^{-1}(A))$$

- Easy for discrete RV: A is countable, $g^{-1}(A)$ is countable $P_y(Y = y) = \sum_{x \in g^{-1}(y)} P_x(X = x)$

- Continuous RV:

  $F_Y(y_0) = \Pr.(Y \leq y_0) = \Pr.(g(X) \leq y_0)$

  $= \int_{\{g(x) \leq y_0\}} p(x)dx$ can be hard to compute for general g !

  Easiest if g(x) is monotone increasing: $g(x) \leq y_0 \Leftrightarrow x \leq g^{-1}(y_0)$

- Result: if X has CDF $F_X(x)$, then the CDF of $Y = g(X)$ is

  If g is an increasing function on D_x: $F_Y(y_0) = F_X(g^{-1}(y_0))$ [1]

  If g is a decreasing function on D_x: $F_Y(y_0) = 1 - F_X(g^{-1}(y_0))$

- We most often want the PDF of y: use the chain rule on [1]

$$p_Y(y_0) = p_X(g^{-1}(y_0)) \left| \frac{d}{dy} g^{-1}(y_0) \right|$$

The term in the absolute value is the Jacobian of the inverse transformation

Applies to the multivariate case.

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \xrightarrow[g]{} \begin{cases} y_1 = g(x_1, x_2) \\ y_2 = g(x_1, x_2) \end{cases}$$

$$\begin{pmatrix} \frac{dy_1}{dx_1} & \frac{dy_1}{dx_2} \\ \frac{dy_2}{dx_1} & \frac{dy_2}{dx_2} \end{pmatrix}$$

$$x_1 = g^{-1}(y_1, y_2)$$

$$x_2 = g^{-2}(y_1, y_2)$$

4.2 Transformation with non-monotone functions

Idea: break the domain in sets where the transform is monotone

- Result: If we can partition X in $A_1, A_2, \ldots A_N$,
  with $f_X(x)$ continuous on each A_i, and **$g_i(x)$ monotone in A_i**.

  Then: $$\boldsymbol{p_Y(y) = \sum_1^n p_X\left(g_i^{-1}(y)\right) \left|\frac{d}{dy} g_i^{-1}(y)\right|}$$

- Example: Square of a unit normal $p(x) = \frac{1}{\sqrt{2\Pi}} e^{-\frac{x^2}{2}}$, $y = g(x) = x^2$.

x<0 $g^{-1}(y) = -\sqrt{y}$ x>0 $g^{-1}(y) = \sqrt{y}$

p(y) = $\left|\frac{-1}{2\sqrt{y}}\right| \frac{1}{\sqrt{2\Pi}} e^{-\frac{y}{2}} + \left|\frac{1}{2\sqrt{y}}\right| \frac{1}{\sqrt{2\Pi}} e^{-\frac{y}{2}} = \boldsymbol{\frac{1}{\sqrt{y}} \frac{1}{\sqrt{2\Pi}} e^{-\frac{y}{2}}}$, $0<y<\infty$.

The Chi-Square distribution with 1 degree of freedom !

4.3 Fundamental Application: Probability Inverse Transform

- X has CDF $F_X(x)$, it is continuous and *strictly* increasing (for technical reasons).

Think of the transformation $y = F_X(x)$

Then y is a random variable on (0,1) .... obvious

y is uniformly distributed on (0,1) ... not obvious ! $F_Y(y_0) = P(Y \le y_0) = y_0$?

Proof: $P(Y \le y_0) = P(F_X(X) \le y_0)$

$= P(X \le F^{-1}_X(y_0))$

$= F_X(F^{-1}_X(y_0))$

$= y_0$

- Application: Random number generation by **Inverse Transform**

Want to generate random numbers from X with CDF $F_X(x)$

1. Generate a uniform number u_0 from U~(0,1)

2. Compute x_0: $F_X(x_0) = u_0$. That is, compute $x_0 = F_X^{-1}(u_0)$

.... The **Inverse Transform** Method