# Econ 2110, fall 2016, Part Ic
# Review of Probability Theory

Maximilian Kasy

Department of Economics, Harvard University

# Roadmap

- Ia
  - Basic definitions
  - Conditional probability and independence
- Ib
  - Random Variables
  - Expectations
  - Transformation of variables
- Ic
  - Selected probability distributions
  - Inequalities

# Part Ic

Selected probability distributions

Inequalities

# Selected probability distributions
# Discrete Distributions

**Bernoulli**

- $X$ takes on the values 0 and 1
- $f_X(1) = p$,
  $f_X(0) = 1 - p$.
- Example: will a given person find a job in the next month?

# Binomial

- suppose $X_i$ are iid Bernoulli with parameter $p$
- let $Y = \sum_{i=1}^{n} X_i$
- then $Y$ takes on the values $S = \{0, 1, \cdots, n\}$
- for $y \in S$
$$f_Y(y) = \frac{n!}{(n-y)! \cdot y!} p^y (1-p)^{n-y}$$
- Example: Number of highschool dropouts in a random sample of size $n$, when population share of dropouts is $p$

# Poisson distribution

- $X$ takes on the values $\{0, 1, 2, \cdots\}$
- $$f_X(x) = \frac{m^x e^{-m}}{x!}$$
- $E[X] = Var[X] = m$
- useful for modeling 'successes' that occur over intervals of time (people finding jobs, atoms decaying,...).
- limit as $n \to \infty$ of a Binomial distribution with parameter $p_n = m/n$
- Example: Number of people in the US finding a new job before noon today

# Continuous Distributions

**Uniform distribution**:

- $f_X(x) = \mathbf{1}[a \leq x \leq b](b-a)^{-1}$ for $b > a$.
- Example: $F_X(X)$ is uniform $(0,1)$ distributed for any continuously distributed $X$

# Univariate normal distribution

- **Standard normal**
  $Z \sim \mathcal{N}(0,1)$
  $$f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\tfrac{1}{2} z^2\right]$$
- **General normal**
  $X \sim \mathcal{N}(\mu, \sigma^2)$.
- Let $X = \mu + \sigma Z$ for $\sigma > 0$.
- from the transformation formula we get
  $$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}\right]$$

- $E[X] = \mu$,
  $E[X^2] = \sigma^2 + \mu^2$,
  $E[(X - \mu)^2] = \sigma^2$
  and $E[(X - \mu)^4] = 3\sigma^4$
- Importance of normals:
  averages of independent stuff are approximately normal
- central limit theorem – see part IV of class
- Examples: test scores, asset returns, physical height

# Chi-Squared distribution

- now come several distributions derived from the standard normal
- very useful for constructing confidence sets and tests (Part IV of class)
- we already did linear transformations, how about squares?
- let $X_i \sim iid \mathcal{N}(0,1)$
- let $Y = \sum_{i=1}^{k} X_i^2$
- then $Y$is distributed chi-squared with $k$ degrees of freedom
- $Y \sim \chi_k^2$
- $E[Y] = k$
  and $Var[Y] = 2k$ (Why?)

# F-distribution

- let $Y_1 \sim \chi^2_k$ and $Y_2 \sim \chi^2_l$
- where $Y_1$ and $Y_2$ are independent
- let
$$Q = \frac{Y_1/k}{Y_2/l}$$
- then $Q$ is distributed $F$ with $k$ degrees of freedom in the numerator and $l$ degrees of freedom in the denominator
- $Q \sim F_{k,l}$

# Student's t-distribution

- let $Z \sim \mathcal{N}(0, 1)$, and $Y \sim \chi^2_k$
- where $Z$ and $Y$ are independent
- let
$$T = \frac{Z}{\sqrt{Y/k}}$$
- then $T$ is distributed student-t with $k$ degrees of freedom
- $T \sim t_k$

# Multivariate Normal Distribution

- $X = (X_1, \cdots, X_n)$ has a multivariate normal distribution
- if and only if $\alpha' X$ is normally distributed
- for all $\alpha \in \mathbb{R}^n$.
- This definition allows that $P(\alpha' X = 0) = 1$ for some $\alpha$.

- $X$ multivariate normal $\Rightarrow X_i$ is normally distributed.
- The mean and covariance matrix of $X$ exist.
- Denote them by $\mu$ and $\Sigma$
- Let $\alpha$ be any nonstochatic $n \times 1$ vector. Then $Y = \alpha' X$ is normal with mean and variance

$$E[\alpha' X] = \mu' \alpha \qquad \text{and} \qquad Var[a' X] = a' \Sigma a$$

- let $\beta$ be a $k \times 1$ nonstochastic vector, and let $B$ and be $n \times k$. let $Y = \beta + B' X$. then

$$Y \sim \mathcal{N}(\beta + B' \mu, B' \Sigma B)$$

# Density of multivariate normal

- Can derive it from density of standard normal.
- If $Z \sim \mathscr{N}_n(0, I_n)$, then $Z_i$ are iid standard normal.
- Independence $\Rightarrow$ joint density is product of densities

$$f_Z(z) = \prod_{i=1}^{n} (2\pi)^{-1/2} \exp[-\tfrac{1}{2} z_i^2] = (2\pi)^{-n/2} \exp[-\tfrac{1}{2} z' z].$$

- suppose $\Sigma$ is full rank
  let $X = \Sigma^{1/2} Z + \mu$,
  with inverse transformation $Z = \Sigma^{-1/2}(X - \mu)$
- $\Rightarrow X \sim N(\mu, \Sigma)$
- Jacobian of the transformation: $|\Sigma^{-1/2}| = |\Sigma|^{-1/2}$
- $\Rightarrow$ by the transformation formula

$$f_X(x) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp\left[-\tfrac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right]$$

# Conditional distribution of multivariate normal

Let

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$$

- $X_1$ and $X_2$ are independent if and only if $\Sigma_{12} = 0$.
  (holds only for normals!)
- Suppose $\Sigma_{22}$ is full rank.
- Then the conditional distribution of $X_1$, given $X_2 = x_2$, is given by

$$X_1 | X_2 = x_2 \sim \mathcal{N}\left( \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \right).$$

- The regression function

$$\mu_{X_1}(x_2) = E[X_1 | X_2 = x_2]$$

  is linear in $x_2$.
- This holds for normals, but not in general!

## Proposition

Let $A$ be an $n \times n$ matrix which is

- symmetric: $A' = A$
- idempotent: $A^2 = A$.

If $Z \sim N(0, I)$, then
$$s^2 := Z'AZ \sim \chi^2_p,$$
where $p$ is the trace of $A$.

**Proof:**

- Symmetric matrices have orthonormal eigenvectors $P$
- Eigenvalue decomposition:

$$A = P \Lambda P'$$

- Let $\widetilde{Z} = P'Z$. Then $\operatorname{Var}(\widetilde{Z}) = PP' = I$,

$$\widetilde{Z} \sim N(0, I)$$

  and

$$s^2 = Z'AZ = \widetilde{Z}' \Lambda \widetilde{Z} = \sum \lambda_i \widetilde{Z}_i^2 .$$

- Idempotent matrices have eigenvalues $\lambda_i$ equal to 0 or 1.
- $\operatorname{tr}(A) = \sum \lambda_i$.

## Proposition (Distribution of t-statistic)

- Suppose $X_i \sim iid \mathcal{N}(\mu, \sigma^2)$.
- Let $\overline{x} = \frac{1}{n} e' X$, where $e = (1, \ldots, 1)$
- Let $s^2 = \frac{1}{n-1} \sum (X_i - \overline{x})^2$
- Then

$$\frac{\sqrt{n}(\overline{x} - \mu)}{\sqrt{s^2}} \sim t_{n-1}$$

**Proof:**

- the claim follows if we can show that
  1. $\sqrt{n}(\overline{x}-\mu) \sim N(0,\sigma^2)$
  2. $\frac{1}{\sigma^2}s^2 \sim \chi^2_{n-1}$
  3. $\overline{x}$ and $s^2$ are independent
- 1 is easy
- to show 2, rewrite
$$s^2 = \frac{1}{n-1}X'MX$$
where
$$M = I - \frac{1}{n}ee'$$
is symmetric, idempotent, and has trace $n-1$
- to show 3, let $Y = MX$, so that $s^2 = \frac{1}{n-1}Y'Y$,
note that $\overline{x}$ and $Y$ are jointly normally distributed, and
$$\mathrm{Cov}(\overline{x}, Y) = \sigma^2 \frac{1}{n} e'M = 0.$$

# Inequalities

- often too hard / cumbersome to compute some properties of random variables
- easier to bound these properties
- useful especially in asymptotics (part IV of class)
- allows to show that we can neglect some remainder terms in large samples, etc.

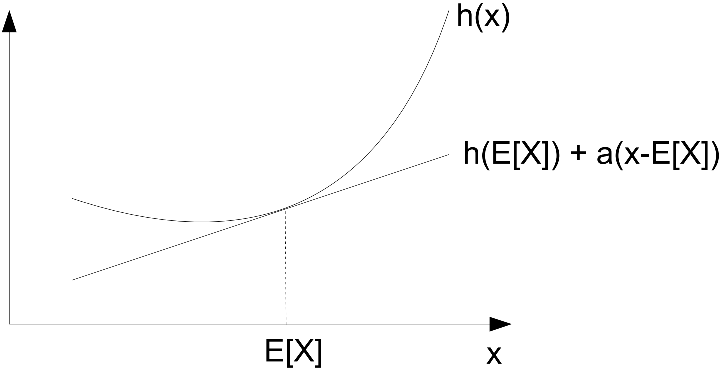
# Jensen's inequality

## Proposition

- Let $h(x)$ be a convex function $h : \mathbb{R} \mapsto \mathbb{R}$.
- Let $X$ be a random variable.
- Then

$$E[h(X)] \geq h(E[X]).$$

Figure: Proof of Jensen's inequality



**Proof:**

- convexity $\Rightarrow$ there is an $a$ such that

$$h(x) \geq h(E[X]) + a(X - E[X])$$

- take expectations on both sides

$$E[h(x)] \geq h(E[X]) + a(E[X] - E[X]) = h(E[X]).$$

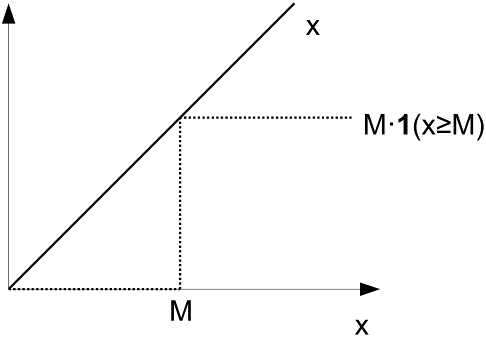
# Markov's inequality

## Proposition

- Suppose $X$ is a random variable,
- $X \geq 0$, and $E[X] < \infty$.
- Then, for all $M > 0$

$$P(X \geq M) \leq \frac{E[X]}{M},$$

Figure: Proof of Markov's inequality



**Proof:**

- $X \geq 0 \Rightarrow$

$$X \geq M \cdot \mathbf{1}(X \geq M)$$

- Take expectations on both sides $\Rightarrow$

$$E[X] \geq M \cdot P(X \geq M).$$

# Chebychev Inequality

## Proposition

- Suppose $X$ is a random variable,
- such that $\sigma^2 = Var[X] < \infty$.
- Then, for all $M > 0$

$$P(|X - \mu| \geq M) \leq \frac{\sigma^2}{M^2}$$

where $\mu = E[X]$.

**Proof:**

- Let $Y = (X - \mu)^2$
- Apply Markov's inequality to $Y$ and the cutoff $M^2$

$$P(Y \geq M^2) \leq \frac{E[Y]}{M^2}$$

- Rewrite

$$P(|X - \mu| \geq M) \leq \frac{\sigma^2}{M^2}$$