Advanced Econometric Methods EMET3011/8014

Lecture 4

John Stachurski

Semester 1, 2011

Announcements

- Please get a fresh copy of the course notes PDF
- Errors? Typos? Unclear? Please visit/email me

Today's Lecture

- Probability Models
- Random Variables
- Expectations
- Distributions
- Dependence

Probability Theory

The foundation of stats, and hence econometrics

Probability theory is built on measure theory

Measure theory outside the scope of this course

Some statements in following are missing technical caveats

For the purposes of this course, we can ignore exceptions

Probability Theory

The modeling of random outcomes

Start with **sample space**

- set of all possible outcomes in a random experiment
- can be any nonempty set
- typically denoted Ω

A typical element of Ω is denoted ω

Let $\mathcal{F}$ denote set of all subsets of Ω

- Elements of $\mathcal{F}$ are also called **events**

For example, $\emptyset \in \mathcal{F}$ and $\Omega \in \mathcal{F}$

Example

Let $\Omega := \{1, \ldots, 6\}$ represent the six different faces of a dice.

A typical element of $\mathcal{F}$ is

$$\{2, 4, 6\} = \{ \text{ all even faces } \}$$

Probabilities

In advanced probability theory, we first assign probability to events

(Not individual outcomes—see course notes)

Let $A \in \mathcal{F}$

$\mathbb{P}(A)$ represents the “probability that event A occurs.”

“Event A occurs” means:

when $\omega \in \Omega$ is selected by “nature,” $\omega \in A$

We want $\mathbb{P}$ to satisfy some axioms. . .

A **probability** $\mathbb{P}$ on $\mathcal{F}$ is a function from $\mathcal{F} \to [0,1]$ that satisfies

1. $\mathbb{P}(\Omega) = 1$, and
2. If A and B are disjoint events, then

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$$

Second property is called **additivity**

Note: Some technical details omitted. See course notes.

Claim. If $\mathbb{P}$ is a probability and $A_1, \ldots, A_J$ are disjoint, then

$$\mathbb{P}\left(\cup_{j=1}^{J} A_j\right) = \sum_{j=1}^{J} \mathbb{P}(A_j)$$

Proof for disjoint A, B, C:

$A \cup B \cup C = (A \cup B) \cup C$, so

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}((A \cup B) \cup C)$$

Since $A \cup B$ and C are disjoint,

$$\mathbb{P}((A \cup B) \cup C) = \mathbb{P}(A \cup B) + \mathbb{P}(C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C)$$

Example

Let $\Omega := \{1, \ldots, 6\}$

Let $\mathbb{P}(A) := \#A/6$ for $A \in \mathcal{F}$, where $\#A :=$ num elements in A

Additivity holds: given disjoint A and B, we have

$$\begin{aligned}\mathbb{P}(A \cup B) &= \#(A \cup B)/6 \\ &= (\#A + \#B)/6 = \#A/6 + \#B/6 = \mathbb{P}(A) + \mathbb{P}(B)\end{aligned}$$

Additivity is a natural requirement on $\mathbb{P}$:

$$\mathbb{P}\{2,4,6\} = \mathbb{P}[\{2\} \cup \{4\} \cup \{6\}] = \mathbb{P}\{2\} + \mathbb{P}\{4\} + \mathbb{P}\{6\}$$

Example

Memory chip is made up of billions of tiny switches/bits

- Switches can be off or on (zero or 1)

Random number generator accesses N bits, switching each one on or off

We take

- $\Omega := \{(b_1, \ldots, b_N) : \text{where } b_n \text{ is 0 or 1 for each } n\}$
- $\mathbb{P}(A) := 2^{-N}(\#A)$

Exercise: Show that $\mathbb{P}$ is a probability

Fact. If $\mathbb{P}$ is a probability on $\mathcal{F}$ and $A, B \in \mathcal{F}$ with $A \subset B$, then

1. $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$
2. $\mathbb{P}(A) \leq \mathbb{P}(B)$
3. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
4. $\mathbb{P}(\emptyset) = 0$

Proof.

When $A \subset B$, we have $B = (B \setminus A) \cup A$ and hence

$$\mathbb{P}(B) = \mathbb{P}(B \setminus A) + \mathbb{P}(A)$$

All results follow (why?) □

Fact. If A and B are any events, then

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

Implication: For any $A, B \in \mathcal{F}$, we have

$$\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$$

Exercise: Check the fact. Hint: $A = [(A \cup B) \setminus B] \cup (A \cap B)$

Conditional probability of A given B is

$$\mathbb{P}(A \mid B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \tag{1}$$

Probability of A, given information that B has occurred

Events A and B called **independent** if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$

- If A and B independent, then

$$\mathbb{P}(A \mid B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A)$$

Example

Experiment: roll a dice twice.

$$\Omega := \{(i,j) : i,j \in \{1,\ldots,6\}\} \quad \text{and} \quad \mathbb{P}(E) := \#E/36$$

Now consider the events

$$A := \{(i,j) \in \Omega : i \text{ is even}\} \quad \text{and} \quad B := \{(i,j) \in \Omega : j \text{ is even}\}$$

In this case we have

$$A \cap B = \{(i,j) \in \Omega : i \text{ and } j \text{ are even}\}$$

Exercise: Verify that $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$

Hence, A and B are independent under the probability $\mathbb{P}$

Law of total probability:

If $A \in \mathcal{F}$ and $B_1, \ldots, B_M$ is a partition of Ω with $\mathbb{P}(B_m) > 0$ for all m, then

$$\mathbb{P}(A) = \sum_{m=1}^{M} \mathbb{P}(A \mid B_m) \cdot \mathbb{P}(B_m)$$

Proof.

Given any such A and partition $B_1, \ldots, B_M$, we have

$$\begin{aligned} \mathbb{P}(A) = \mathbb{P}[A \cap (\cup_{m=1}^{M} B_m)] &= \mathbb{P}[\cup_{m=1}^{M} (A \cap B_m)] \\ &= \sum_{m=1}^{M} \mathbb{P}(A \cap B_m) = \sum_{m=1}^{M} \mathbb{P}(A \mid B_m) \cdot \mathbb{P}(B_m) \end{aligned}$$

$\square$

Random variables

Informally: A "value that changes randomly"

Formally: A **random variable** x is a function from Ω into $\mathbb{R}$

Interpretation: random variables convert outcomes in sample space into numerical outcomes

General idea:

- "nature" picks out ω in Ω
- random variable now reports outcome as $x(\omega) \in \mathbb{R}$

Note: Some technical details omitted. See course notes.

Example

Suppose Ω is set of infinite binary sequences

$$\Omega := \{(b_1, b_2, \ldots) : b_n \in \{0,1\} \text{ for each } n\}$$

We can create different random variables mapping $\Omega \to \mathbb{R}$:

- Number of "flips" till first "heads":

$$x(\omega) = x(b_1, b_2, \ldots) = \min\{n : b_n = 1\}$$

- Number of "heads" in first 10 "flips":

$$x(\omega) = x(b_1, b_2, \ldots) = \sum_{n=1}^{10} b_n$$

Common notational convention with RVs:

$$\{x \text{ has some property}\} := \{\omega \in \Omega : x(\omega) \text{ has some property}\}$$

Example:

$$\{x \leq 2\} := \{\omega \in \Omega : x(\omega) \leq 2\}$$

$$\therefore \quad \mathbb{P}\{x \leq 2\} := \mathbb{P}\{\omega \in \Omega : x(\omega) \leq 2\}$$

Example: Given random variable x and $a \leq b$, we claim that

$$\mathbb{P}\{x \leq a\} \leq \mathbb{P}\{x \leq b\} \tag{2}$$

This holds because

$$\begin{aligned} \{x \leq a\} := \{\omega \in \Omega : x(\omega) \leq a\} \\ \subset \{\omega \in \Omega : x(\omega) \leq b\} := \{x \leq b\} \end{aligned}$$

Now apply monotonicity: $A \subset B \implies \mathbb{P}(A) \leq \mathbb{P}(B)$

Discrete random variables

Let Q be a statement, such as "a is greater than 3"

Definition: $\mathbb{1}\{Q\}$ equals one if Q true, zero otherwise

Bernoulli random variable:

$$x(\omega) = \mathbb{1}\{\omega \in C\} \quad \text{where } C \in \mathcal{F}$$

- x is a binary indicator of whether or not C occurs

A **discrete random variable** is an RV with finite range:

$$x \text{ discrete} \iff \# \operatorname{rng}(x) < \infty$$

Discrete RVs can be formed by taking "linear combinations" of Bernoulli RVs

Example:

$$x(\omega) = s\,\mathbb{1}\{\omega \in A\} + t\,\mathbb{1}\{\omega \in B\}$$

Visualization when $\Omega = \mathbb{R}$ (not generally true):

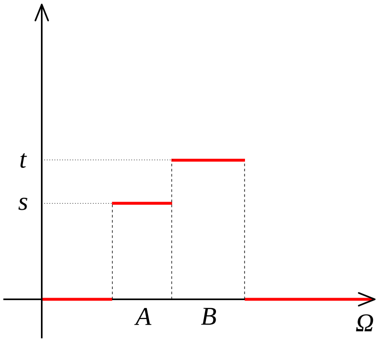


Figure: $x(\omega) = s\,\mathbb{1}\{\omega \in A\} + t\,\mathbb{1}\{\omega \in B\}$

A general expression for a discrete RV:

$$x(\omega) = \sum_{j=1}^{J} s_j \mathbb{1}\{\omega \in A_j\} \tag{3}$$

- the s_j's are distinct
- the A_j's are a partition of Ω

Convince yourself:

- $x(\omega) = s_j$ if and only if $\omega \in A_j$
- $\{x = s_j\} = A_j$
- $\mathbb{P}\{x = s_j\} = \mathbb{P}(A_j)$

Expectations

We want to define expectations for an arbitrary RV x

Roughly speaking, $\mathbb{E}[x] :=$ the "sum" of all possible values of x, weighted by their probabilities.

"Sum" in quotes because may be infinite number of possibilities

Definition: For the discrete RV

$$x(\omega) = \sum_{j=1}^{J} s_j \mathbb{1}\{\omega \in A_j\} \tag{4}$$

the **expectation** $\mathbb{E}[x]$ of x is given by

$$\mathbb{E}[x] := \sum_{j=1}^{J} s_j \mathbb{P}(A_j) \tag{5}$$

Intuition: Since $A_j = \{x = s_j\}$ we have

$$\mathbb{E}[x] = \sum_{j=1}^{J} s_j \mathbb{P}\{x = s_j\}$$

Consequence: Expectation of an indicator equals prob of the event

$$\mathbb{E}\left[\mathbb{1}\{\omega \in A\}\right] = \mathbb{P}(A) \quad \text{for all } A \in \mathcal{F}$$

Proof.

We have

$$\mathbb{1}\{\omega \in A\} = 1 \times \mathbb{1}\{\omega \in A\} + 0 \times \mathbb{1}\{\omega \in A^c\}$$

Applying the definition,

$$\mathbb{E}\left[\mathbb{1}\{\omega \in A\}\right] = 1 \times \mathbb{P}(A) + 0 \times \mathbb{P}(A^c) = \mathbb{P}(A)$$

$\square$

Fact. The expectation of a constant α is α

True meaning: α is the constant random variable $\alpha \mathbb{1}\{\omega \in \Omega\}$

Proof.

From the definition,

$$\mathbb{E}[\alpha] := \mathbb{E}[\alpha \mathbb{1}\{\omega \in \Omega\}] = \alpha \mathbb{P}(\Omega) = \alpha$$

$\square$

How about the expectation of an arbitrary random variable x?

The full definition involves measure theory, and we omit it

Short story: any x can be approximated by a sequence of finite-valued random variables x_n.

The expectation of x is then defined as

$$\mathbb{E}[x] := \lim_{n \to \infty} \mathbb{E}[x_n]$$

When things are done carefully (details omitted), this value doesn't depend on the particular approximating sequence $\{x_n\}$

Fact. If x and y are RVs with $x \leq y$, then $\mathbb{E}[x] \leq \mathbb{E}[y]$

Comments:

- This property called **monotonicity of expectations**
- $x \leq y$ means that $x(\omega) \leq y(\omega)$ for all $\omega \in \Omega$

Exercise: Check for $x(\omega) := \mathbb{1}\{\omega \in A\}$ and $y(\omega) := \mathbb{1}\{\omega \in B\}$

Hint: What does $x \leq y$ imply about A and B?

Fact. If x and y are RVs and α and β are constants, then

$$\mathbb{E}\left[\alpha x + \beta y\right] = \alpha \mathbb{E}\left[x\right] + \beta \mathbb{E}\left[y\right]$$

- This property called **linearity of expectations**
- $(\alpha x + \beta y)(\omega) := \alpha x(\omega) + \beta y(\omega)$

We just show that $\mathbb{E}[\alpha x] = \alpha \mathbb{E}[x]$ for $x(\omega) := \sum_{j=1}^{J} s_j \mathbb{1}\{\omega \in A_j\}$

Let $y := \alpha x$

$$y(\omega) = \alpha x(\omega) = \alpha \left[\sum_{j=1}^{J} s_j \mathbb{1}\{\omega \in A_j\} \right] = \sum_{j=1}^{J} \alpha s_j \mathbb{1}\{\omega \in A_j\}$$

Hence, by the def of expectations for discrete RVs

$$\mathbb{E}[y] = \sum_{j=1}^{J} \alpha s_j \mathbb{P}(A_j) = \alpha \left[\sum_{j=1}^{J} s_j \mathbb{P}(A_j) \right] = \alpha \mathbb{E}[x]$$

Variance and Covariance

The ***k*-th moment of *x*** is defined as $\mathbb{E}\left[x^k\right]$ for $k \in \mathbb{N}$

The **variance** of x is defined as

$$\operatorname{var}[x] := \mathbb{E}\left[(x - \mathbb{E}[x])^2\right]$$

The **standard deviation** of x is $\sqrt{\operatorname{var}[x]}$

- Measure the dispersion of x

The **covariance** of random variables x and y is defined as

$$\operatorname{cov}[x, y] := \mathbb{E}\left[(x - \mathbb{E}[x])(y - \mathbb{E}[y])\right]$$

Fact. If α and β are constants and x and y are random variables, then

- $\operatorname{var}[\alpha] = 0$
- $\operatorname{var}[\alpha + \beta x] = \beta^2 \operatorname{var}[x]$
- $\operatorname{var}[\alpha x + \beta y] = \alpha^2 \operatorname{var}[x] + \beta^2 \operatorname{var}[y] + 2\alpha\beta \operatorname{cov}[x, y]$

Exercise: Check all these facts

Given RVs x and y with variances σ_x^2 and σ_y^2,

$$\textbf{correlation} \text{ of } x \text{ and } y \;:= \operatorname{corr}[x,y] := \frac{\operatorname{cov}[x,y]}{\sigma_x \, \sigma_y}$$

If $\operatorname{corr}[x,y] = 0$, we say that x and y are **uncorrelated**

Fact. Given RVs x and y, constants $\alpha, \beta > 0$, we have

$$-1 \leq \operatorname{corr}[x,y] \leq 1 \quad \text{and} \quad \operatorname{corr}[\alpha x, \beta y] = \operatorname{corr}[x,y]$$

CDFs

A **cumulative distribution function** (cdf) on $\mathbb{R}$ is a function $F\colon \mathbb{R} \to [0,1]$ that is

- right-continuous,
- monotone increasing, and satisfies
- $\lim_{s\to-\infty} F(s) = 0$ and $\lim_{s\to\infty} F(s) = 1$

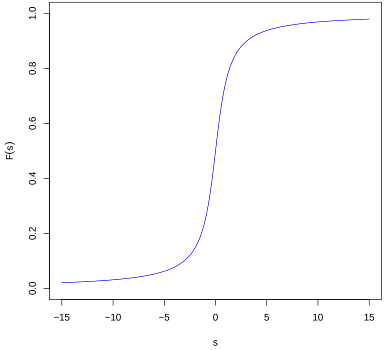
Definitions:

F is right continuous if $s_n \downarrow s$ implies $F(s_n) \downarrow F(s)$

F is monotone increasing if $s \leq s'$ implies $F(s) \leq F(s')$

Example

The function $F(s) = \arctan(s)/\pi + 1/2$ is a cdf



Given RV x, the **distribution** F_x of x is the function

$$F_x(s) := \mathbb{P}\{x \leq s\} \qquad (s \in \mathbb{R})$$

Theorem. F_x is a cdf for any RV x

For example, F_x is monotone increasing, because if $s \leq s'$, then

$$F_x(s) := \mathbb{P}\{x \leq s\} \leq \mathbb{P}\{x \leq s'\} =: F_x(s')$$

(Further details omitted—see course notes for related exercises)

Notation: $x \sim F$ means x has distribution F

Exercise: Show that $1 - F(s) = \mathbb{P}\{x > s\}$

Fact. If $x \sim F$ and $a \leq b$, then $\mathbb{P}\{a < x \leq b\} = F(b) - F(a)$

Proof.

Recall that $A \subset B$ implies $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$

If $a \leq b$, then

$$\{a < x \leq b\} = \{x \leq b\} \setminus \{x \leq a\} \quad \text{and} \quad \{x \leq a\} \subset \{x \leq b\}$$

$$\therefore \quad \mathbb{P}\{a < x \leq b\} = \mathbb{P}\{x \leq b\} - \mathbb{P}\{x \leq a\}$$

$\square$

Cdf F called **symmetric** if $F(-s) = 1 - F(s)$ for all $s \in \mathbb{R}$

- Probability that $x \leq -s$ equals prob that $x > s$

Fact. Let $x \sim F$. If

- F is symmetric
- $\mathbb{P}\{x = s\} = 0$ for all $s \in \mathbb{R}$

then the cdf $F_{|x|}$ of $|x|$ is given by

$$F_{|x|}(s) := \mathbb{P}\{|x| \leq s\} = 2F(s) - 1 \qquad (s \geq 0)$$

Proof: An exercise

Densities and Probability Mass Functions

Two convenient special cases

- Discrete: cdf is just jumps (a step function)
- Density: cdf smooth with no jumps

A **probability mass function** (pmf) is a finite sequence $p_1, \ldots, p_J$ with

- $0 \leq p_j \leq 1$ for each j
- $\sum_{j=1}^{J} p_j = 1$

A **density** is a function $p \colon \mathbb{R} \to \mathbb{R}$ with

- $p(s) \geq 0$ for all s
- $\int_{-\infty}^{\infty} p(s) ds = 1$

Discrete Case. Suppose that x takes values $s_1, \ldots, s_J$

Letting $p_j := \mathbb{P}\{x = s_j\}$, the cdf corresponding to x is

$$F_x(s) = \sum_{j=1}^{J} \mathbb{1}\{s_j \leq s\} p_j \qquad (s \in \mathbb{R})$$

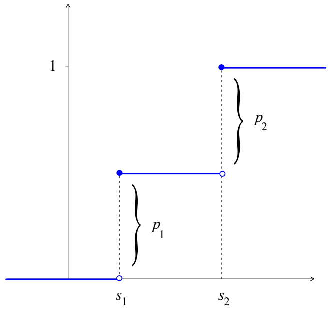
because

$$F_x(s) := \mathbb{P}\{x \leq s\} = \mathbb{P} \bigcup_{j \text{ s.t. } s_j \leq s} \{x = s_j\}$$

$$= \sum_{j \text{ s.t. } s_j \leq s} \mathbb{P}\{x = s_j\} = \sum_{j=1}^{J} \mathbb{1}\{s_j \leq s\} p_j$$

Exercise: Show that $p_1, \ldots, p_J$ form a pmf

Example: $x = s_i$ with probability p_i



Density Case.

Suppose cdf F smooth (derivative F' exists), let $p := F'$

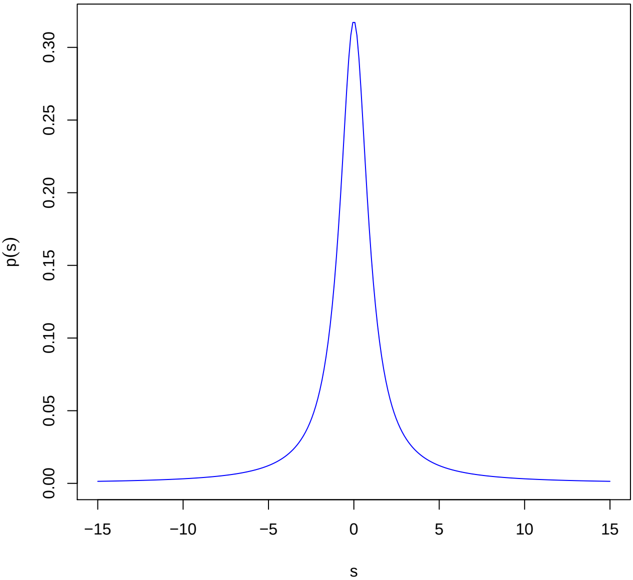
By the fundamental theorem of calculus,

$$\int_r^s p(t)dt = F(s) - F(r) \quad \text{for any} \quad r \leq s$$

From the definition of cdfs, we can see that

- $p(s) \geq 0$ for all s
- $\int_{-\infty}^{+\infty} p(s)ds = 1$

In other words, p is a density


0.00
0.05
0.10
0.15
0.20
0.25
0.30
p(s)
−15
−10
−5
0
5
10
15
s

In general: If $x \sim F$ and $p = F'$, then p is called the **density of x**

Note that

- cdf is fully specified by density because $F(s) = \int_{-\infty}^{s} p(t)dt$
- not every random variable has a density

Fact. If x has a density, then $\mathbb{P}\{x = s\} = 0$ for all $s \in \mathbb{R}$

The Quantile Function

If F a strictly increasing function, then inverse function F^{-1} exists:

$$\text{the function } F^{-1} \text{ such that } F^{-1}(F(s)) = s$$

For general F set

$$F^{-1}(q) := \inf\{s \in \mathbb{R} : F(s) \geq q\} \qquad (0 < q < 1)$$

The function F^{-1} is called the **quantile function**

- $F^{-1}(0.25)$ called the first quartile of F
- $F^{-1}(0.5)$ is called the **median** of F
- $F^{-1}(0.75)$ called the third quartile of F

Quantile function used in hypothesis testing

Example: Let $x \sim F$, where F is

- strictly increasing
- differentiable
- symmetric

Let $\alpha \in (0,1)$

Find c such that $\mathbb{P}\{-c \leq x \leq c\} = 1 - \alpha$

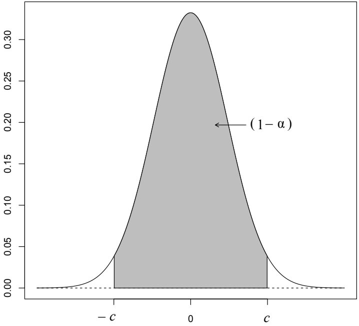


Figure: Finding critical values

Claim: $c = F^{-1}(1 - \alpha/2)$ implies $\mathbb{P}\{-c \leq x \leq c\} = 1 - \alpha$

Proof.

We have

$$\begin{aligned}
\mathbb{P}\{-c \leq x \leq c\} &= \mathbb{P}[\{x = -c\} \cup \{-c < x \leq c\}] \\
&= \mathbb{P}\{x = -c\} + \mathbb{P}\{-c < x \leq c\} \\
&= \mathbb{P}\{-c < x \leq c\} \\
&= F(c) - F(-c)
\end{aligned}$$

By symmetry,

$$F(c) - F(-c) = F(c) - 1 + F(c) = 2F(c) - 1$$

Since $F(c) = 1 - \alpha/2$, we have $2F(c) - 1 = 1 - \alpha$ $\square$

Distributions in R

R has functions for accessing all common distributions.

Functions have the form `lettername()`

- "letter" is one of `p` , `d`, `q` or `r`
- "name" is one of the named distributions in R

The meanings of the letters are

`p`	cumulative distribution function
`d`	density function
`q`	quantile function
`r`	generates random variables

Examples:

```
> pnorm(2, mean=.1, sd=2)  # F(2), F = cdf of N(.1, 4)
> qcauchy(1/2)  # median of cauchy distribution
> runif(100, 2, 4)  # 100 uniform r.v.s on [2, 4]
```

See documentation for further details

Expectations from Distributions

Expectations can be calculated from distributions

Let $h \colon \mathbb{R} \to \mathbb{R}$.

If x has density p, then

$$\mathbb{E}\left[h(x)\right] = \int_{-\infty}^{\infty} h(s)p(s)ds$$

If x is a discrete RV with $\mathbb{P}\{x = s_j\} = p_j$, then

$$\mathbb{E}\left[h(x)\right] = \sum_{j=1}^{J} h(s_j)p_j$$

Proof of second case is in the course notes

Convenient notation to unify:

$$\text{If } x \sim F \text{, we write } \mathbb{E}\left[h(x)\right] = \int h(s)F(ds)$$

Meaning:

- In density case, $\int h(s)F(ds) := \int_{-\infty}^{\infty} h(s)p(s)ds$
- If discrete case, $\int h(s)F(ds) := \sum_{j=1}^{J} h(s_j)p_j$

Note: $\int h(s)F(ds)$ is actually the L-S integral—see course notes

Dependence

Consider N random variables $x_1, \ldots, x_N$, where $x_n \sim F_n$

F_n tells us about properties of x_n viewed as a single entity

How about the relationships between the the variables $x_1, \ldots, x_N$?

To quantify, we define the **joint distribution** of $x_1, \ldots, x_N$ to be

$$F(s_1, \ldots, s_N) := \mathbb{P}\{x_1 \leq s_1, \ldots, x_N \leq s_N\}$$

In this setting, F_n sometimes called the **marginal distribution**

Joint density (if exists) is a function p satisfying

$$\int_{-\infty}^{t_N} \cdots \int_{-\infty}^{t_1} p(s_1, \ldots, s_N) ds_1 \cdots ds_N = F(t_1, \ldots, t_N)$$

for all $t_n \in \mathbb{R}$, $n = 1, \ldots, N$.

Conditional density of $x_{k+1}, \ldots, x_N$ given $x_1 = s_1, \ldots, x_k = s_k$ is defined by

$$p(s_{k+1}, \ldots, s_N \,|\, s_1, \ldots, s_k) := \frac{p(s_1, \ldots, s_N)}{p(s_1, \ldots, s_k)}$$

Rearranging gives decomposition

$$p(s_1, \ldots, s_N) = p(s_{k+1}, \ldots, s_N \,|\, s_1, \ldots, s_k) p(s_1, \ldots, s_k)$$

Independence

Typically, joint distribution cannot be determined from marginals

Once special case where we can:

RVs $x_1, \ldots, x_N$ called **independent** if, given any $s_1, \ldots, s_N$, we have

$$F(s_1, \ldots, s_N) = F_1(s_1) \times \cdots \times F_N(s_N)$$

If all marginals the same, then x_n's called **identically distributed**

If also independent, then called IID

Fact. If $x_1, \ldots, x_N$ independent, then

$$\mathbb{E}\left[\prod_{n=1}^{N} x_n\right] = \prod_{n=1}^{N} \mathbb{E}\left[x_n\right]$$

Fact. If $x_1, \ldots, x_N$ independent with marginal density p_n, then joint density p satisfies

$$p(s_1, \ldots, s_N) = \prod_{n=1}^{N} p_n(s_n)$$

Exercise: Show that if x and y are independent, then $\operatorname{cov}[x, y] = 0$

(Converse is not generally true)

(Informal) homework:

- Minimum: Work through exercises in this lecture
- Extra: Exercises in chapter 3 of course notes

Problems? Questions? Come and see one of us.

- JS: Monday 8:30–10:00
- VW: Wed 2:00–3:30
- Other times okay too, just email me