SESSION 1
REVIEW OF PROBABILITY

Define $S \equiv$ the sample space, the set of all possible outcomes of an experiment

"experiment" should be thought of broadly
cf. data-generating process (DGP)
could be discrete, could be ctn.

Let $\mathcal{A}$ be a family of sets, each a subset of S.
$\mathcal{A}$ is a σ-field iff:

1. $S \in \mathcal{A}$
2. if $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$
3. if $A_1, A_2, \ldots$ is an infinite sequence $\in \mathcal{A}$, then
$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$$

Sets in $\mathcal{A}$ are called events.
If S is finite or countable, then $\mathcal{A}$ includes all subsets of
If S is $\mathbb{R}^n$, $\mathcal{A}$ is the σ-field of Borel sets, which is just the smallest σ-field containing all n-dim intervals

2 coinflip ex.

Def. a probability distribution P on $(S, \mathcal{A})$ is a non-negative fn. defined for each event in $\mathcal{A}$ w/ the following:

1. $P(S) = 1$
2. if $A_1, A_2, \ldots$ is a sequence of disjoint events
$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

continue ex.

$\frac{1}{4}$	$\frac{1}{4}$
$\frac{1}{4}$	$\frac{1}{4}$

.7	.05
.05	.2

$(S, \mathcal{A}, P)$ is called a probability space.

Some things to prove about properties of P:

(a) If $A_1, \ldots, A_n$ are disjoint events, then $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$

(b) $P(\phi) = 0$

(c) $P(A^c) = 1 - P(A)$

(d) If $A \subset B$ are two events, then $P(A) \leq P(B)$

(e) If A_1, A_2, ~~and~~ events, then $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 A_2$

(f) If $A_1 \subset A_2 \subset \ldots$ is a sequence of events, then $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n \to \infty} P(A$

Linda is 31, single, outspoken, and very bright. She majored in philosophy in college. As a student she was deeply concerned with racial discrimination and other social issues, and participated in anti-nuclear demonstrations.

Rank the likelihood of various alternatives

1. Linda is active in the feminist movement.
2. Linda is a bank teller.
3. Linda is a bank teller and active in the feminist movemen

Conditional Probability (discrete)

Conditional probability of $A | B \equiv \Pr(A|B) \equiv \dfrac{\Pr(AB)}{P(B)}$ if $P(B) > 0$

Ex. 2 coins in my pocket, one fair, one 2-headed
choose 1 at random and flip, it's heads
what is the probability it was ~~the~~ fair coin?

On your own: same question, but
(b) second flip also heads
(c) 3rd flip is tails

Bayes' theorem

perhaps the most important one of all

Let $A_1, A_2, \ldots$ be an infinite sequence of disjoint events s.t.
$\bigcup_{i=1}^{\infty} A_i = S$ and $\Pr(A_i) > 0 \ \forall i$, and let B be an event w/ $\Pr(B)$:

Then
$$\underbrace{\Pr(A_i | B)}_{\text{posterior probabilities}} = \frac{\Pr(B|A_i)\Pr(A_i)}{\sum_{j=1}^{\infty} \Pr(B|A_j)\underbrace{\Pr(A_j)}_{\text{sometimes called prior probabilities}}}$$

Pf. Fix i.
By def of conditional prob, $\Pr(A_i|B) = \dfrac{\Pr(A_i B)}{\Pr(B)} = \dfrac{\Pr(BA_i)}{\Pr(B)}$

and also $\Pr(BA_i) = \Pr(B|A_i)\Pr(A_i)$

By the properties of the sequence, $B = \bigcup_{j=1}^{\infty} BA_j$

Since these events are disjoint, $\Pr(B) = \sum_{j=1}^{\infty} \Pr(BA_j) = \sum_{j=1}^{\infty} \Pr(B|A_j)\Pr$ QED

Bayes thm example

1. revisit the fair vs. 2-headed coin example

2. A disease hits 1 in 10,000 people.
   There's a test that's 90% accurate.
   You take the test and it comes back positive.
   What's the cond'l prob you have the disease?

independence

Def A and B are indep. if $Pr(AB) = Pr(A)Pr(B)$

Ex. (easy) 2-coin flip

Ex. (harder) throw a fair 6-sided die

event A : even number

event B : 1, 2, 3, or 4

Are A & B indep?

Intuition: if info on B is of no help in predicting A, indep!

anything about it

Measurable functions

Given S and $\mathcal{A}$, let g be a fn. defined at all points of $S \to \mathbb{R}$.
$\forall x \in \mathbb{R}$, let A_x be the subset of S defined by

$$A_x = \{ s : g(s) \leq x \}$$

Then g is measurable w.r.t. $\mathcal{A}$, or $\mathcal{A}$-measurable, if $A_x \in \mathcal{A}$ $\forall x \in \mathbb{R}$.

Intuition: think of measurability w/r/t an information set.
It's measurable if it's well defined based on the current granularity of $\mathcal{A}$
Ex. 2-coin sequential flip; after 1 flip, fn of second flip

Random Variables
Given a prob. space $(S, \mathcal{A}, P)$, a random variable is an $\mathcal{A}$-measurable fn w/value $X(s)$ $\forall s \in S$

Since X is a $f: S \to \mathbb{R}$, it induces a probability dist'n on $\mathbb{R}$
That is, for any Borel set $B \subset \mathbb{R}$

$$P_X(B) = Pr(X \in B) = Pr\{ s : X(s) \in B \}$$

Distribution function (or cumulative d.f. "CDF"): a fn. F s.t. $F(t) = Pr(X \leq t)$
Properties of F:

non decreasing

rt. continuous

$\lim_{t \to -\infty} F(t) = 0$ $\qquad \lim_{t \to \infty} F(t) = 1$

Discrete dist'ns

have a probability mass fn (PMF) $f(x) = \Pr(X = x)$

and so. $$P_X(B) = \sum_{i:\, x_i \in B} f(x_i)$$

Abs. cts. distns

r.v. X has an abs. cts. distn if $\exists$ non-neg. prob. density fn (PDF) f s.t. $\forall$ Borel sets $B \subset \mathbb{R}$,
$$P_X(B) = \int_B f(x)\, dx$$

Properties: CDF can be differentiated almost everywhere $F'(x) = f(x)$

m-variate distns are just the prob. distn of (2 or more m) r.v.'s.

joint dist'n fn or joint CDF defined as
$$F(x_1, \ldots, x_m) = \Pr(X_1 \le x_1, \ldots, X_m \le x_m)$$

For abs cts distns, the pdf is $f(x_1, \ldots, x_m)$ s.t. $\forall$ Borel $B \subset \mathbb{R}^m$
$$\Pr[(X_1, \ldots, X_m) \in B] = \int \cdots \int_B f(x_1, \ldots, x_m)\, dx_1 \cdots dx_m$$

Measure theory unites ~~the~~ discrete & abs. cts. distns w/ notation

probability measure μ

$$\int_B g(x) f(x)\, d\mu(x)$$ for Borel B in $\mathbb{R}$ or sometimes $\int_B g(x)\, dF(x)$

For discrete distns, this ~~...~~ means
$$\sum_{x \in B} g(x)\, f(x) \quad \leftarrow \text{PMF}$$

For abs cts distns
$$\int_B g(x)\, f(x)\, dx \quad \leftarrow \text{pdf}$$

Quantile functions Let F be a CDF

Def. p quantile is $\min x : F(x) \ge p$.

often written as $x = F^{-1}(p)$

ex. VaR

$p = \frac{1}{2}$ median

$p = \frac{1}{4}$ lower quartile

$p = \frac{3}{4}$ upper quartile

marginal distn – joint dist'n of a subset of the m-tuple $X_1, \ldots, X_m$
let $1 \le k < m$. Then

CDF
$$F_k(x_1, \ldots, x_k) = Pr(X_1 \le x_1, \ldots, X_k \le x_k)$$
$$= \lim_{x_{k+1}, \ldots, x_m \to \infty} F_m(x_1, \ldots, x_m)$$

PDF
$$f_k(x_1, \ldots, x_k) = \int \cdots \int_{R^{m-k}} f_m(x_1, \ldots, x_m)\, d\mu(x_{k+1}) \cdots d\mu(x_m)$$

independence

$X_1, \ldots, X_m$ are independent iff $\exists\, G_1 \ldots G_m$ s.t.
$\forall (x_1, \ldots, x_m) \in \mathbb{R}^m$, $F_m(x_1, \ldots, x_m) = G_1(x_1) \cdots G_m(x_m)$

A similar factorization holds for the pdf

random vectors: An n-dim. random vector $\underset{\sim}{X}$ is
just a sequence of n random variables $\underset{\sim}{X}' = (X_1, \ldots, X_n)$

ex. for marginals
$$f(x,y) = \begin{cases} cx^2y & x^2 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$
what are the marginals? what is c?

Moments

if it exists, The 1^{st} moment or expectation of any r.v. X is given by

$$E(X) = \int_R x \, dF(x)$$

More generally, if g is an integrable fn of X, then

$$E[g(X)] = \int_R g(x) \, dF(x)$$

ex. $X \sim U(-1,1)$

$E(X^2) = ?$

$E[X^r]$ is the r^{th} moment of X, if it exists.

$E[(X - E(X))^r]$ is the r^{th} central moment

Variance of X = $Var(X) = E[(X - E(X))^2]$

$$= E(X^2) - [E(X)]^2$$

For any 2 r.v's X and Y:

Covariance(X,Y) = $Cov(X,Y) = E[(X - E(X))(Y - E(Y))]$

$$= EXY - EXEY$$

For random vectors,

$$E(\underset{\sim}{X}) = [E(X_1), \ldots, E(X_n)]'$$

$$Cov(\underset{\sim}{X}) = E[(\underset{\sim}{X} - E(\underset{\sim}{X}))(\underset{\sim}{X} - E(\underset{\sim}{X}))'] = \begin{bmatrix} Var(X_1) & Cov(X_1,X_2) & \cdots & Cov(X_1,X_n) \\ & \ddots & & \\ & & & Var(X_n) \end{bmatrix}$$

Def: $\underset{\sim}{Y} = \underset{m\times n}{A} \underset{n\times 1}{\underset{\sim}{X}} + \underset{n\times 1}{\underset{\sim}{b}}$ $\qquad E(\underset{\sim}{X}) = \underset{n\times 1}{\underset{\sim}{\mu}}$

Then $E(\underset{\sim}{Y}) = A\underset{\sim}{\mu} + \underset{\sim}{b}$ $\qquad Cov(Y) = A \Sigma A'$

Characteristic Functions / Moment Generating Functions

The characteristic fn of a r.v. X is a fn. $\phi(t) = E(e^{itX})$
$= E(\cos tX) + i\,E(\sin tX)$

Properties:

The char fn. always exists, is defined $\forall\, t \in \mathbb{R}$

one-to-one correspondence btw. CDFs and char fns.

$X_1, \dots, X_k$ indep. w/ char fns. $\phi_1, \dots, \phi_k$, then
$Y = X_1 + \dots + X_k$ has char fn. $\phi_Y(t) = \prod_{j=1}^{k} \phi_j(t)$

Main useful property

if $E(X^r)$ exists, then $E(X^r) = i^{-r} \left.\dfrac{d^r\phi_X(t)}{dt^r}\right|_{t=0}$

we'll return to this when we talk about the normal dist'n

Char fns easily extend to m-variate random vectors $\underset{\sim}{X}$: $(X_1, \dots, X_m)$

$$\underset{m\times 1}{\phi(\underset{\sim}{t})} = E\Big(e^{i t' \underset{m\times 1}{\underset{\sim}{X}}}\Big)$$

this can be used to find moments and cross-moments:

$$E\left(\prod_{j=1}^{n} X_j^{r_j}\right) = \frac{1}{i^{r_1+\dots+r_n}}\left[\frac{\partial^{r_1+\dots+r_n}\phi(\underset{\sim}{t})}{\partial t_1^{r_1}\cdots\partial t_n^{r_n}}\right]_{\underset{\sim}{t}=\underset{\sim}{0}}$$

Moment generating fns. are very similar. For ~~an~~ r.v. X, MGF
$\psi(t) = E(e^{tX})$

Unlike the char fn, the MGF may not exist, but if it does, we can use it to recover moments:

$\psi'(0) = E(X) \qquad \psi''(0) = E(X^2) \qquad \psi'''(0) = E(X^3) \quad \dots$

Other props:

1. Sums of indep. r.v.'s has MGF that is the products of MGF's

2. $Y = aX + b$. $\forall t$ s.t. $\psi_X(at)$ exists, $\psi_Y(t) = e^{bt}\psi_X(at)$

Pf.

$$\psi_Y(t) = E(e^{tY}) = E\left(e^{t(aX+b)}\right) = e^{bt}E(e^{atX}) = e^{bt}\psi_X(at)$$

Ex. Bernoulli distn

$$X = \begin{cases} 1 & \text{w/prob } p \\ 0 & \text{w/prob } q = 1-p \end{cases}$$

$$\text{MGF } \psi_X(t) = E(e^{tX}) = pe^t + (1-p)\cdot 1 = pe^t + q$$

with 1st 2 moments

$$E(X) = \psi'(0) = p$$

$$E(X^2) = \psi''(0) = p \Rightarrow Var(X) = p - p^2 = p(1-p) = pq$$

just as easy to do these moments directly

Now suppose $Y = X_1 + \dots + X_n$, $X_i \overset{iid}{\sim} \text{Bernoulli}(p)$

Then by property 1 above

$$\psi_Y(t) = (pe^t + q)^n$$

and

$$E(Y) = np \qquad var(Y) = np(1-p)$$

Y has a binomial distn, with

$$f(x) = \begin{cases} \binom{n}{x} p^x q^{n-x} & \text{for } x = 0, 1, \dots, n \\ 0 & \text{otherwise} \end{cases}$$

Transformations of r.v.'s

Let f be the joint pdf of abs cts r.v.'s $X_1, \dots, X_n$
Suppose we want the joint pdf of r.v.'s $Y_1, \dots, Y_n$, where
$Y_1 = g_1(\underset{\sim}{X}), \dots, Y_n = g_n(\underset{\sim}{X})$

Let the xformation be one-to-one for some $R^0 \subset \mathbb{R}^n$ s.t. $Pr(\underset{\sim}{x} \in \mathbb{R}^0) = 1$
(this assumption can be weakened considerably, cf. DeGroot (1970))

Then we can write the inverse
$x_1 = h_1(y_1, \dots, y_n) \quad \dots \quad x_n = h_n(y_1, \dots, y_n)$

Assume 1st partial derivatives of all h exist and are cts.
Define the Jacobian $J(y_1, \dots, y_n)$ as the determinant

$$\begin{vmatrix} \frac{\partial h_1}{\partial y_1} & \cdots & \frac{\partial h_1}{\partial y_n} \\ \vdots & & \\ \frac{\partial h_n}{\partial y_1} & \cdots & \frac{\partial h_n}{\partial y_n} \end{vmatrix}$$

If the Jacobian does not vanish at any $(y_1, \dots, y_n)$
then the pdf of Y is given by

$$f(\underline{y}) = f\left[h_1(\underline{y}), \dots, h_n(\underline{y})\right] \left| J(\underline{y}) \right|$$

ex. $f(x_1, x_2) = \begin{cases} 4x_1x_2 & x_1, x_2 \in [0,1] \\ 0 & \text{o.w.} \end{cases}$

$$Y_1 = \frac{X_1}{X_2} \qquad Y_2 = X_1 X_2$$

Solve for the inverse: $X_1 = (Y_1 Y_2)^{1/2} \qquad X_2 = \left(\frac{Y_2}{Y_1}\right)^{\frac{1}{2}}$

Jacobian $\begin{vmatrix} \frac{1}{2}\left(\frac{y_2}{y_1}\right)^{1/2} & \frac{1}{2}\left(\frac{y_1}{y_2}\right)^{1/2} \\ -\frac{1}{2}\left(\frac{y_2}{y_1^3}\right)^{\frac{1}{2}} & \frac{1}{2}\left(\frac{1}{y_1 y_2}\right)^{1/2} \end{vmatrix} = \frac{1}{2y_1}$

Note the support for Y_1 and Y_2

$Y_1 > 0$, $Y_2 > 0$, $Y_1 Y_2 < 1$, $\frac{Y_2}{Y_1} < 1$

So $\left|\frac{1}{2y_1}\right| = \frac{1}{2y_1}$

and $f(y_1, y_2) = \begin{cases} 2\left(\frac{y_2}{y_1}\right) & \text{(in the support)} \\ 0 & \text{otherwise} \end{cases}$

Conditional distributions

Let $X = (X_1, \dots, X_m)'$ and $Y = (Y_1, \dots, Y_n)'$ be 2 random vectors with joint pdf $f(\underset{\sim}{x}, \underset{\sim}{y})$ and marginal $g(\underset{\sim}{y})$ for Y.

Then $\forall \underset{\sim}{y}$ s.t. $g(y) > 0$, the conditional distn $X \mid Y = y$ has pdf

$h(\underset{\sim}{x} \mid \underset{\sim}{y}) = \frac{f(\underset{\sim}{x}, y)}{g(y)}$ or equivalently $f(\underset{\sim}{x}, y) = h(x|y)\, g(y)$

Note that this event could have $\Pr = 0$, so this extends the earlier result on conditioning. Same extension for

Bayes Thm

As above, but define $h_1(x|y)$, $h_2(y|x)$ as conditional pdf's and $g_1(x)$ and $g_2(y)$ as marginals. $\forall \underset{\sim}{x}$ s.t. $g_1(\underset{\sim}{x}) > 0$

$$h_2(y|x) = \frac{h_1(x|y)\, g_2(y)}{\int_{\mathbb{R}^n} h_1(x|t)\, g_2(t)\, d\mu(t)}$$

Conditional Expectations

Again define $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ as above; let $\phi(X,Y)$ be an integrable fn of $\underset{\sim}{X}$ and $\underset{\sim}{Y}$.

Then the cond'l expectation of $\phi(X,Y)$ given Y is a fn. of y

$$E\left[\phi(X,Y) \mid y\right] = \int_{\mathbb{R}^m} \phi(x,y)\, h_1(x \mid y)\, d\mu(x)$$

Law of iterated expectations:

$$E\left[E\left(\phi(X,Y) \mid Y\right)\right] = E\left[\phi(X,Y)\right]$$

Finally, if X and Y are scalars w/ $Var(X) < \infty$, then

$$Var(X) = E\left[Var(X \mid Y)\right] + Var\left[E(X \mid Y)\right]$$

Pf. Exercise

Ex. $X \sim U[0,1]$. Then cond'l on X, $Y \sim U[X,1]$

what's the marginal dist'n of Y?

Marginal pdf of X is $g_1(x) = \begin{cases} 1 & x \in [0,1] \\ 0 & \text{otherwise} \end{cases}$

Cond'l of Y is $h_2(y \mid x) = \begin{cases} \frac{1}{1-x} & x < y < 1 \\ 0 & \text{otherwise} \end{cases}$

Joint pdf is $f(x,y) = \begin{cases} \frac{1}{1-x} & 0 < x < y < 1 \\ 0 & \text{otherwise} \end{cases}$

Marginal of y is $g_2(y) = \int f(x,y)\,dx = \int_0^y \frac{1}{1-x}\,dx =$

$$= \begin{cases} -\ln(1-y) & y \in [0, \\ 0 & \text{otherw} \end{cases}$$