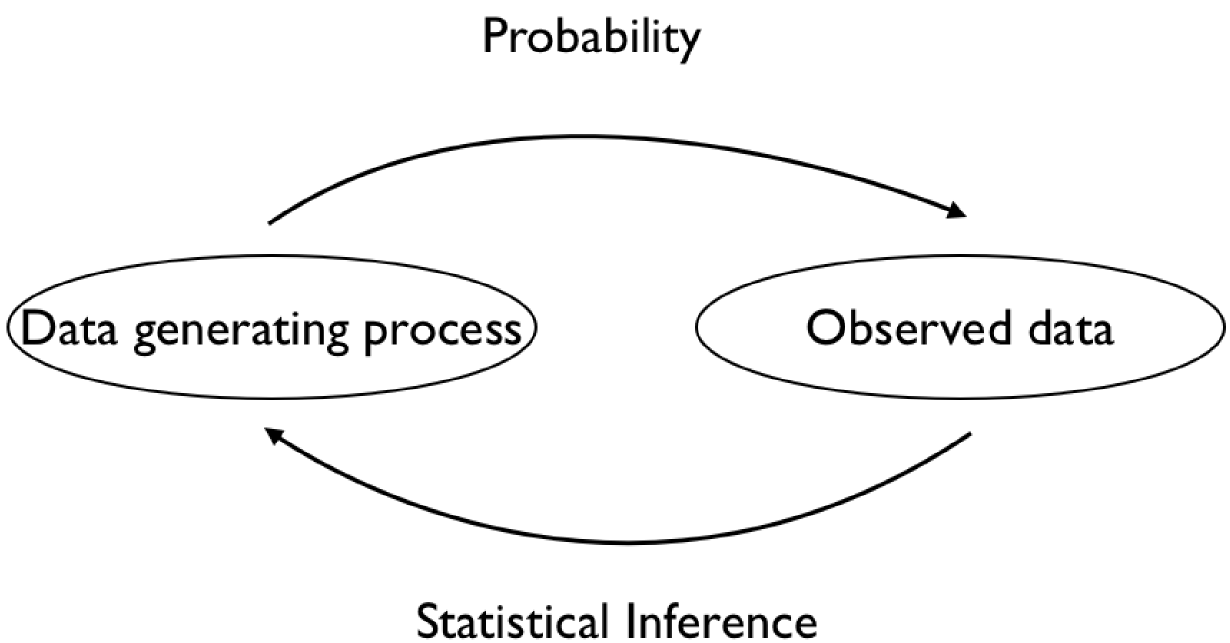
Statistics 200B: Introduction to Probability and Statistics at an Advanced Level

Soyeon Ahn <iamyeoni@berkeley.edu>
Department of Statistics

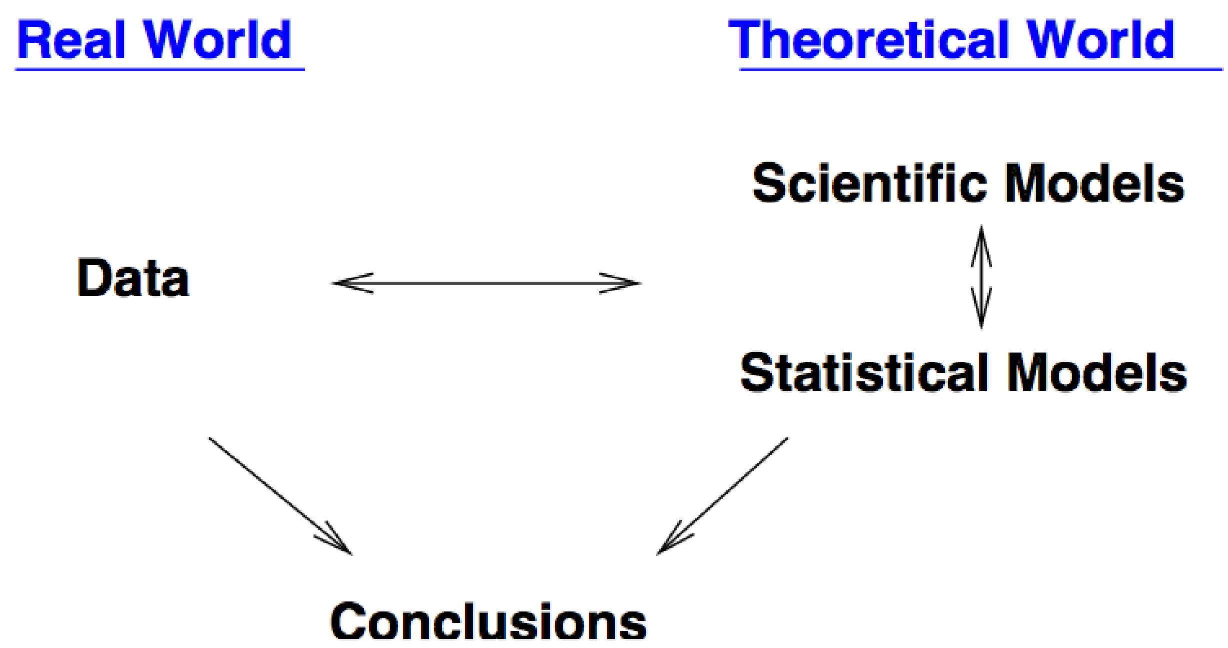
UC Berkeley, Spring 2019

The Big Picture



adapted from Wasserman, 2004

A Slightly Bigger Picture



from Kass, 2009

Probability Review

CDF $F_X : \mathbb{R} \to [0,1]$, $F_X(x) = P(X \leq x)$

Properties:

1. Nondecreasing: $x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2)$
2. Normalized: $\lim_{x\to-\infty} = 0$ and $\lim_{x\to\infty} = 1$
3. Right-continuous: $\lim_{y\downarrow x} F(y) = F(x)$

We write $X \sim F$ to denote that r.v. X has distribution F.

We write $X \stackrel{D}{=} Y$ to denote that X and Y are equal in distribution. This means that $F_X(x) = F_Y(x)\,\forall x$, *not* that $X = Y$.

X is *continuous* if its CDF F_X is continuous.

X is *discrete* if it takes countably many values.

There are random variables that are neither (e.g., F_X is discontinuous at a single point, assigning it non-zero probability), but they can usually be derived from these two main types.

The *density* of a r.v. X has a different definition, depending on whether X is discrete or continuous.

If X is discrete, its density or probability mass function (PMF)

$$f_X(x) = P(X = x)$$

When X is continuous, $P(X = x) = 0$ for all x.

When X is continuous, we define the probability density function (PDF) in terms of integration. It satisfies

1. $f_X(x) \geq 0 \, \forall x$
2. $\int_{-\infty}^{\infty} f_X(x)dx = 1$
3. $P(a < X < b) = \int_a^b f_X(x)dx$

We can go back and forth between the PDF and CDF:

$$
\begin{aligned}
F_X(x) &= \int_{-\infty}^{x} f_X(t)dt \\
f_X(x) &= F'_X(x) \, \forall x \text{ s.t. } F_X(x) \text{ is differentiable}
\end{aligned}
$$

Example:

Let X be a continuous random variable with PDF f. Let A be a subset of the real line, and let $I_A(x)$ be the indicator function for A:

$$I_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

Let $Y = I_A(X)$. Find an expression for the CDF of Y.

Hint: First find the PDF for Y.

Bivariate distributions

Joint CDF $F(x, y) = P(X \leq x, Y \leq y)$

When X and Y are discrete, their joint PMF is $f(x, y) = P(X = x, Y = y)$.

When X and Y are continuous, their joint PDF is $f(x, y)$, satisfying

1. $f(x, y) \geq 0 \, \forall (x, y)$
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$
3. $P((X, Y) \in A) = \int \int_A f(x, y) dx dy$

We can compute *marginal* densities from the joint PDF.

When X is discrete, $f_X(x) = \sum_y f(x, y)$.

When X is continuous, $f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$.

X and Y are *independent* if $f_{X,Y}(x, y) = f_X(x) f_Y(y)$

Obvious analogues hold for more than two variables. We write $X_1, \ldots, X_n \overset{iid}{\sim} F$ ($iid =$ independent and identically distributed) to denote that a collection of random variables are mutually independent with common distribution F. Then we may write

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_1}(x_i)$$

Define the *conditional density*

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

When X and Y are discrete, $P(X = x|Y = y) = f_{X|Y}(x|y)$.

When X and Y are continuous, $P(X \in A|Y = y) = \int_A f_{X|Y}(x|y)dx$.

Bayes Theorem:

$$f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y)f_Y(y)}{f_X(x)}$$

Transformations

Consider a new r.v. $Y = r(X)$. We want to calculate the PDF and CDF of Y based on the PDF and CDF of X.

If X is discrete, it's easier to work directly with the PDFs

$$
\begin{aligned}
f_Y(y) &= P(Y = y) = P(r(X) = y) \\
&= P(\{x : r(x) = y\}) \\
&= P(X \in r^{-1}(y)) \\
&= \sum_{x \in r^{-1}(y)} f(x)
\end{aligned}
$$

Then calculate the CDF based on the PDF.

If X is continuous, it's easier to reason about the CDFs

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(r(X) \leq y) \\ &= P(\{x : r(x) \leq y\}) \\ &= \int_{A_y} f_X(x) dx \end{aligned}$$

where $A_y = \{x : r(x) \leq y\}$.

Then calculate the PDF based on the CDF.

In the special case that r is strictly monotone, with $s = r^{-1}$,

$$f_Y(y) = f_X(s(y)) \left| \frac{ds(y)}{dy} \right|$$

Example

Let X be a r.v. with PDF $f_X(x) = e^{-x}$ for $x > 0$. Let $Y = r(X) = \log X$.

1. Calculate the CDF of X.
2. Calculate the CDF of Y.
3. Calculate the PDF of Y by differentiating the CDF.
4. Calculate the PDF of Y directly from the PDF of X.

Example

Let $X, Y \stackrel{iid}{\sim} Unif(0,1)$. That is X and Y are independent and both have pdf

$$f(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & otherwise \end{cases}$$

Let $Z = \max\{X, Y\}$. Find the density of Z.

Hint: First calculate the CDF of Z, followed by its PDF.

Expectation, variance, and covariance

$$E[X] = \int x dF(x) = \begin{cases} \sum_x x f(x) & X \text{ discrete} \\ \int x f(x) dx & X \text{ continuous} \end{cases}$$

When $E[X] = \mu$,

$$V[X] = E[(X-\mu)^2] = \int (x-\mu)^2 dF(x)$$

When $E[X] = \mu_X$ and $E[Y] = \mu_Y$,

$$Cov(X, Y) = E[(X-\mu_X)(Y-\mu_Y)]$$

Some properties

1. For r.v. $X_1, \ldots, X_n$ and constants $a_1, \ldots, a_n$,

$$
\begin{aligned}
E\left(\sum_i a_i X_i\right) &= \sum_i a_i E[X_i] \\
V\left(\sum_i a_i X_i\right) &= \sum_i a_i^2 V(X_i) + 2 \sum \sum_{i<j} a_i a_j Cov(X_i, X_j)
\end{aligned}
$$

2. $V[X] = E[X^2] - (EX)^2$; $Cov(X, Y) = E(XY) - E(X)E(Y)$

3. If X and Y are independent, $E[XY] = E[X]E[Y]$ and $Cov(X, Y) = 0$.

"Law of the Lazy Statistician": If $Y = r(X)$, then

$$E[Y] = \int r(x) dF_X(x)$$

Special case: $E[I_A(x)] = \int I_A(x) dF(x) = \int_A dF(x) = P(X \in A)$

Note $\int_A dF(x) = \begin{cases} \sum_{x \in A} f(x) & X \text{ discrete} \\ \int_A f(x) dx & X \text{ continuous} \end{cases}$

Don't confuse the mean and variance (which I'll often denote by μ and σ^2) with the *sample mean* and *sample variance*.

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \qquad S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

Conditional expectation

$$E(X|Y=y) = \begin{cases} \sum_x x f_{X|Y}(x|y) & \text{discrete case} \\ \int x f_{X|Y}(x|y)dx & \text{continuous case} \end{cases}$$

Note! $E(X|Y=y)$ is a function of y. Therefore $E(X|Y)$ is a r.v.

Iterated expectations and variances

$$\begin{aligned} E[Y] &= E[E[Y|X]] \\ V[Y] &= E[V[Y|X]] + V[E[Y|X]] \end{aligned}$$

Moment generating function (MGF) $\phi_X(t) = E[e^{tX}] = \int e^{tX} dF(x)$

$$E[X^k] = \phi^{(k)}(0)$$

Markov's inequality: Let X be a non-negative r.v. with finite mean. Then

$$P(X > t) \leq \frac{E[X]}{t} \text{ for any } t > 0$$

Chebyshev's inequality: Let $E[X] = \mu$ and $V[X] = \sigma^2$. Then

$$P(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2} \text{ for any } t > 0$$

Cauchy-Schwartz inequality: If X and Y have finite variances, then

$$E|XY| \leq \sqrt{E[X^2]E[Y^2]}$$

Jensen's inequality: If g is convex, then $E[g(X)] \geq g(E[X])$

Convergence of random variables

Let $X_1, X_2, \ldots$ be a sequence of r.v.'s and let X be another r.v. Let F_n be the CDF of X_n and F be the CDF of X.

X_n converges to X in probability, written $X_n \stackrel{P}{\to} X$, if, for every $\epsilon > 0$,

$$P(|X_n - X| > \epsilon) \to 0$$

X_n converges to X in distribution, written $X_n \stackrel{D}{\to} X$, if

$$\lim_{n\to\infty} F_n(t) = F(t) \text{ at all } t \text{ for which } F \text{ is continuous}$$

Sometimes I'll abuse notation and put F on the right hand side of the arrow, e.g. $X_n \stackrel{D}{\to} N(0,1)$ instead of $X_n \stackrel{D}{\to} Z$ where $Z \sim N(0,1)$.

Example

Suppose $Y_1, Y_2, \ldots$ are iid $Unif(0,1)$ r.v. and let $X_n = \max_{1 \leq i \leq n} X_i$.

1. Show that $X_n \xrightarrow{P} 1$.

2. Show that the r.v. $n(1 - X_n) \xrightarrow{D} Exponential(1)$.

X_n converges to X in quadratic mean (also called convergence in mean square or convergence in L_2), written $X_n \stackrel{qm}{\to} X$, if

$$E[(X_n - X)^2] \to 0 \text{ as } n \to \infty$$

$X_n \stackrel{qm}{\to} b$ if and only if $E[X_n] \to b$ and $V[X_n] \to 0$.

X_n converges to X almost surely, written $X_n \stackrel{as}{\to} X$, if

$$P(X_n \to X \text{ as } n \to \infty) = 1$$

Example

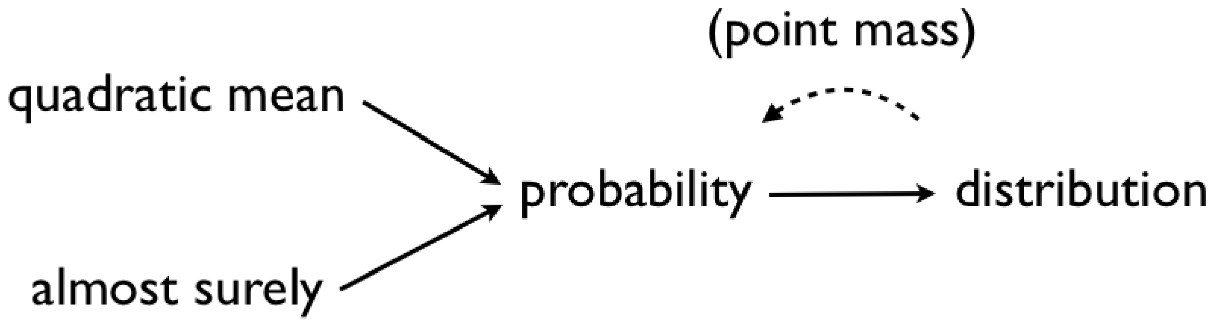
Let $X_1, X_2, \ldots$ be a sequence of r.v.'s such that

$$P\left(X_n = \frac{1}{n}\right) = 1 - \frac{1}{n^2} \quad \text{and} \quad P(X_n = n) = \frac{1}{n^2}$$

1. Does X_n converge in probability?

2. Does X_n converge in quadratic mean?

Relationships



1. $X_n \stackrel{qm}{\to} X$ implies that $X_n \stackrel{P}{\to} X$

2. $X_n \stackrel{as}{\to} X$ implies that $X_n \stackrel{P}{\to} X$

3. $X_n \stackrel{P}{\to} X$ implies that $X_n \stackrel{D}{\to} X$

4. $X_n \stackrel{D}{\to} c$ (point mass) implies that $X_n \stackrel{P}{\to} c$

Slutsky's Theorem

Let X_n, X, and Y_n be random variables and c be a constant. If $X_n \stackrel{D}{\to} X$ and $Y_n \stackrel{D}{\to} c$, then

1. $X_n + Y_n \stackrel{D}{\to} X + c$

2. $X_n Y_n \stackrel{D}{\to} cX$

Questions to consider: What types of convergence are preserved under addition? Under multiplication? Under continuous mappings?

The Weak Law of Large Numbers

Let $X_1, X_2, \ldots$ be an *iid* sample, with $\mu = E[X_1]$ and $\sigma^2 = V[X_1] < \infty$. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then $\bar{X}_n \stackrel{P}{\to} \mu$ as $n \to \infty$.

Flash forward: In the context of statistics, this property (convergence in probability of a sequence of estimators to a parameter) is known as *consistency*.

The Strong Law of Large Numbers

Under the same conditions above, $\bar{X}_n \stackrel{as}{\to} \mu$ as $n \to \infty$.

The Central Limit Theorem

Let $X_1, X_2, \ldots$ be an *iid* sample, with $\mu = E[X_1]$ and $\sigma^2 = V[X_1] < \infty$. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then

$$Z_n \equiv \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{D} N(0,1)$$

We may also write $Z_n \approx N(0,1)$, $\bar{X}_n \approx N(\mu, \sigma^2/n)$, etc.

Note that this result still holds if we replace σ by S_n, where

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

The Delta Method

Suppose that

$$\frac{\sqrt{n}(Y_n - \mu)}{\sigma} \xrightarrow{D} N(0,1)$$

and that g is a differentiable function s.t. $g'(\mu) \neq 0$. Then

$$\frac{\sqrt{n}[g(Y_n) - g(\mu)]}{|g'(\mu)|\sigma} \xrightarrow{D} N(0,1)$$

In other words,

$$Y_n \approx N\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{implies that} \quad g(Y_n) \approx N\left(g(\mu), [g'(\mu)]^2 \frac{\sigma^2}{n}\right)$$

Example

Let $X_1, X_2, \ldots$ be an *iid* sample, with $\mu = E[X_1]$ and $\sigma^2 = V[X_1] < \infty$. Use the Delta Method to find a function of $X_1, \ldots, X_n$ (also known as a statistic) to approximate the variance of $1/\bar{X}_n$.