Given Ω is chosen, a probability law on Ω is a mapping $\mathbb{P}$ that assigns a number to every event (ie to every subset of Ω) such that:

1) $\mathbb{P}(A) \geq 0$ for every event A
2) $\mathbb{P}(\Omega) = 1$ (normalization)
3) Additivity Rules
   i) If $A \cap B = \emptyset$ (A,B are events) then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$
   ii) If $A_1, A_2, A_3, \ldots$ is a countable sequence of mutually disjoint events (i.e. $A_i \cap A_j = \emptyset \ \forall i \neq j$), then $\mathbb{P}(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$

$\bigcup_{n=1}^{\infty} A_n$ — union of all of them

So, given an event $A \subset \Omega$, $\mathbb{P}(A)$ is a model for "the likelihood that the outcome of the uncertain experiment is in A"

"Event A occurs" means "outcome of experiment is in A"

Conditional Expected Values

Given X, A

$$\mathbb{E}(X|A) = \int_{-\infty}^{+\infty} x f_{X|A}(x)\,dx$$

Given X, Y

$$\mathbb{E}(X|\{Y=y\}) = \int_{-\infty}^{+\infty} x f_{X|Y}(x|y)\,dx \quad \forall y$$

Expected Value Rule

$$\mathbb{E}(g(X)|A) = \int_{-\infty}^{+\infty} g(x) f_{X|A}(x)\,dx$$

$$\mathbb{E}(g(X)|Y=y) = \int_{-\infty}^{+\infty} g(x) f_{X|Y}(x|y)\,dx \quad \forall y$$

Recall the Total Probability-type results

- If events $A_1, A_2, \ldots, A_n$ have >0 probability and partition Ω then
  - $f_X(x) = \sum_{k=1}^{n} f_{X|A_k}(x)\,\mathbb{P}(A_k)$
  - $f_X(x) = \int_{-\infty}^{+\infty} f_{X|Y}(x|y) f_Y(y)\,dy$

From these follows

Total Expectation Theorems

$$\mathbb{E}(X) = \sum_{k=1}^{n} \mathbb{E}(X|A_k)\,\mathbb{P}(A_k)$$

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} \mathbb{E}(X|Y=y) f_Y(y)\,dy$$

Product Rule: If $D_1 \supset D_2 \supset D_3 \supset \cdots \supset D_n$ is a nested decreasing sequence of positive-probability events, then

$$\mathbb{P}(D_n) = \mathbb{P}(D_1)\mathbb{P}(D_2|D_1)\mathbb{P}(D_3|D_2)\cdots\mathbb{P}(D_n|D_{n-1})$$

If $A_1, A_2, \ldots, A_n$ is a partition of Ω, then $C_1, C_2, \ldots, C_n$ partitions B, where

$$C_k = B \cap A_k \quad 1 \leq k \leq n$$

Idea: $B = B \cap \Omega = B \cap (A_1 \cup \cdots \cup A_n) = (B \cap A_1) \cup (B \cap A_2) \cup \cdots \cup (B \cap A_n)$

By additivity,

LAW OF TOTAL PROBABILITY

$$\mathbb{P}(B) = \mathbb{P}(C_1) + \cdots + \mathbb{P}(C_n)$$
$$\mathbb{P}(B) = \mathbb{P}(B \cap A_1) + \cdots + \mathbb{P}(B \cap A_n)$$
$$\boxed{\mathbb{P}(B) = \mathbb{P}(B|A_1)\mathbb{P}(A_1) + \mathbb{P}(B|A_2)\mathbb{P}(A_2) + \cdots + \mathbb{P}(B|A_n)\mathbb{P}(A_n)}$$

Given Ω, $\mathbb{P}$, if $A_1, \ldots, A_n$ are events that partition Ω and have nonzero probability, then for any event B, and for any $k: 1 \leq k \leq n$,

$$\mathbb{P}(A_k|B) = \frac{\mathbb{P}(B|A_k)\mathbb{P}(A_k)}{\mathbb{P}(B|A_1)\mathbb{P}(A_1) + \cdots + \mathbb{P}(B|A_n)\mathbb{P}(A_n)}$$

"PROOF:" Numerator $= \mathbb{P}(B \cap A_k)$ by definition of conditional probability $\quad \frac{\mathbb{P}(A_k \cap B)}{\mathbb{P}(B)} = \mathbb{P}(A_k|B)$!

Denominator $= \mathbb{P}(B)$ by total probability theorem

X, Y independent $\Leftrightarrow F_{X,Y}(x,y) = F_X(x)F_Y(y) \quad \forall x, y$

$\Downarrow$ implies (take $\partial/\partial x$, $\partial/\partial y$)

For X, Y BOTH continuous w/ densities $f_X(x), f_Y(y)$; $f_{X,Y}(x,y)$:

X, Y independent $\Leftrightarrow f_{X,Y}(x,y) = f_X(x) f_Y(y) \quad \forall x, y$

Comment: When X, Y independent, we have

- $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$
- $\mathbb{E}(g(X)h(Y)) = \mathbb{E}(g(X))\mathbb{E}(h(Y))$
- $Var(X+Y) = Var(X) + Var(Y)$

Conditional Independence: $\Omega, \mathbb{P}$; say events A and B are conditionally independent given (event) C when

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

$$\mathbb{P}(A|B \cap C) = \mathbb{P}(A|C); \quad \mathbb{P}(B|A \cap C) = \mathbb{P}(B|C)$$

Knowledge of B gives no functional info about probability of A on top of knowledge of C

To see this just play w/ formulas

$$\mathbb{P}(A|B \cap C) = \frac{\mathbb{P}(A \cap B \cap C)}{\mathbb{P}(B \cap C)} = \frac{\mathbb{P}(A \cap B|C)\,\mathbb{P}(B|C)}{\mathbb{P}(B|C)} = \mathbb{P}(A|C)$$

Given a discrete r.v X w/ $p_X(x)$ pmf, define the expected value (or expectation)

$$\mathbb{E}(X) = \sum_{x \in X} x\, p_X(x)$$

Counting Principle:

Re-count # of subsets of $\Omega = \{s_1, \ldots, s_n\}$.

We can view each subset as arising from a multi stage building process.

Stage 1: Choose whether to put s_1 in the subset -either yes or no $(n_1 = 2)$

Stage 2: Choose whether to put s_2 in the subset -either yes or no $(n_2 = 2)$

$\vdots$

Stage n: Choose whether to put s_n in the subset -either yes or no $(n_n = 2)$

$$\#(\text{subsets}) = n_1 n_2 \cdots n_n = 2^n$$

"n choose k"

$$\binom{n}{k} := \frac{n(n-1)\cdots(n-k+1)}{k!}$$

This comes up in situations involving independent trials

Idea: Perform a random experiment repeatedly + independently

If experiment has 2 outcomes, call them Bernoulli Trials, Generally visualize coin flips in this case - H,T possible outcomes at each stage - if $\mathbb{P}(\{H\}) = p$, $\mathbb{P}(\{T\}) = 1-p$ and you perform experiment n times, for each k, $0 \leq k \leq n$,

$$\mathbb{P}(\text{sequence you get}) = p^k(1-p)^{n-k}, \quad \text{where } k = \#(H) \text{ you flip}$$

This is true REGARDLESS of the order each H,T appears in the sequence of flips.

$$\mathbb{P}(\text{you get } k \text{ H's}) = \#(\text{ways of picking } k \text{ spots in } n \text{ available spots}) = \binom{n}{k} p^k (1-p)^{n-k} \quad 0 \leq k \leq n$$

Next BIG TOPIC: DISCRETE RANDOM VARIABLES

Start with Ω and $\mathbb{P}$; a discrete random variable (r.v.) is a real valued function with domain Ω that takes on only finite or countably infinite number of different values.

i.e. $X: \Omega \to \mathbb{R}$ "X is a mapping from Ω to real's"

p_X is defined as follows: for every possible value of $x \in X$,

$$p_X(x) = \mathbb{P}(A_x) \quad \text{where} \quad A_x = \{s \in \Omega : X(s) = x\}$$

ie $p_X(x) =$ the probability that the r.v. X takes on the specific value x

Book uses $\mathbb{P}(\{X = x\})$ or $\mathbb{P}(X = x)$ ← misuse of notation compared to above

to refer to $\mathbb{P}(A_x)$, where $A_x = \{s \in \Omega : X(s) = x\}$

Things to note about p_X:

- $p_X(x) \geq 0$ for all possible values of X (why? cause for any x, $p_X(x)$ is $\mathbb{P}(\text{an event}) \geq 0$!)
- If V is any finite or countably infinite set of possible values of X, then if we set $B =$ the event "$X \in V$" i.e. $B = \{s \in \Omega : X(s) \in V\}$ then $\mathbb{P}(B) = \sum_{x \in V} p_X(x)$

Strong Law of Large Numbers: When X_n iid, mean μ, common variance σ^2, and $M_n = \frac{X_1 + \cdots + X_n}{n}$, we have

common

$$M_n \xrightarrow{\text{w.p. 1}} \mu \quad \text{as} \quad n \to \infty$$

Expected value rule for joints.

$$\mathbb{E}(g(X,Y)) = \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy\; g(x,y) f_{X,Y}(x,y)$$

Next, multiple discrete rvs and joint pmfs, etc

Given $\Omega, \mathbb{P}$ and two discrete rvs X, Y defined in Ω, define the joint pmf of X and Y via

$$p_{X,Y}(x,y) = \mathbb{P}(\underbrace{\{X=x\}}_{A_x} \cap \underbrace{\{Y=y\}}_{B_y}) \quad \forall x \in X, y \in Y.$$

Note: for any set V of possible value pairs for X, Y, we have

$$\sum_{(x,y) \in V} p_{X,Y}(x,y) = \mathbb{P}(\text{event that } (X,Y) \in V)$$

Since X, Y are discrete rvs, they have their own pmfs p_X, p_Y. These are determined as follows from the joint pmf $p_{X,Y}(x,y)$:

① $p_X(x) = \sum_y p_{X,Y}(x,y) \quad \forall x$

② $p_Y(y) = \sum_x p_{X,Y}(x,y) \quad \forall y$

Why are these true? by Total Probability Thm

$$p_X(x) = \mathbb{P}(\{X = x\}) = \mathbb{P}(A_x) = \sum_y \mathbb{P}(A_x \cap B_y) = \sum_y p_{X,Y}(x,y) \; \forall x$$

These definitions generalize in an obvious way to >2 rvs defined on same $\Omega, \mathbb{P}$.

KEY THING: - joint pmf determines the marginals
- marginals do **NOT** determine the joint

Recall the expected value rule: Given X, $p_X(x)$, $Y = g(X)$, have

$$\mathbb{E}(Y) = \sum_x g(x) p_X(x)$$ ← good because saves us from computing $p_Y(y)$ first before $\mathbb{E}(Y)$

Similarly, given X, Y w/ joint pmf $p_{X,Y}(x,y)$ and some real valued function $Z = g(X,Y)$, we have

$$\mathbb{E}(Z) = \sum_{x \in X} \sum_{y \in Y} g(x,y) p_{X,Y}(x,y)$$ ← don't need $p_Z(z)$ to get $\mathbb{E}(Z)$

Next, Conditional Stuff

Given $\Omega, \mathbb{P}$ and a discrete rv X defined on Ω and an event $A \subseteq \Omega$ w/ $\mathbb{P}(A) > 0$, and a possible value x for X, the conditional pmf of X given A is defined as

$$p_{X|A}(x) = \frac{\mathbb{P}(\{X = x\} \cap A)}{\mathbb{P}(A)} = \text{"}\mathbb{P}(B|A)\text{" where } B \text{ is the event } \{X = x\}\text{"}$$

Observe that for any A w/ $\mathbb{P}(A) > 0$, $p_{X|A}(x)$ as x ranges over X's value space defines a pmf - ie $p_{X|A}(x) \geq 0 \ \forall x$ and $\sum_{x \in X} p_{X|A}(x) = 1$

Here's a fact that's similar to (and follows directly from) the Total Probability Thm: If events $A_1, A_2, \ldots, A_n$ partition Ω, and $\mathbb{P}(A_k) > 0$ for $1 \leq k \leq n$, then for any discrete rv X on Ω,

$$p_X(x) = \sum_{k=1}^{n} p_{X|A_k}\,\mathbb{P}(A_k)$$

More often, encounter conditional pmf of X given some other rv Y (defined on same $\Omega, \mathbb{P}$) - given X, Y defined on $\Omega, \mathbb{P}$, conditional pmf of X given Y is defined for all x and for all y with $\mathbb{P}(\{Y = y\}) = p_Y(y) > 0$ as

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$ ← Same as $p_{X|A}(x)$ where $A = \{Y = y\}$

Standard Notation

Note that for any y w/ $p_Y(y) > 0$, $p_{X|Y}(x|y)$ as x ranges over X values defines a pmf

ie $p_{X|Y}(x|y) \geq 0$ and $\sum_{x \in X} p_{X|Y}(x|y) = 1$

Given X, p_X, and $Y = g(X)$,

$$\mathbb{E}(Y) = \sum_{x \in X} g(x) p_X(x)$$

Conditional Variance

Given X, Y conditional variance of X given Y is the random variable

$$Var(X|Y) = \mathbb{E}((X - \mathbb{E}(X|Y))^2 | Y)$$

A recipe similar to the "g-thing" for computing $Var(X|Y)$

- Given y, compute
  $$Var(X|Y=y) = \mathbb{E}[(X - \mathbb{E}(X|Y=y))^2 | Y=y]$$
  - Do this by finding conditional pmf $p_{X|Y}(x|y)$ or pdf $f_{X|Y}(x|y)$ and then computing variance of it
- This yields a function of y - $\gamma(y)$ - plug Y in for y that yields
  $$Var(X|Y) = \gamma(Y)$$

How to compute? In general

- Find conditional pdf/pmf $f_{X|Y}(x|y)$ / $p_{X|Y}(x|y)$
- Mean of that is $\mathbb{E}(X|Y=y)$
- Variance of that is

$$Var(X - \mathbb{E}(X|Y)|Y=y) = \begin{cases} \int_x (x - \mathbb{E}[X|Y=y])^2 f_{X|Y}(x|y)\,dx \\ \text{OR} \\ \sum_x (x - \mathbb{E}[X|Y=y])^2 p_{X|Y}(x|y) \end{cases}$$

Law of Total Variance: (a sometimes useful identity)

$$Var(X) = \mathbb{E}[Var(X|Y)] + Var[\mathbb{E}(X|Y)]$$

Total probability rules If $A_1, \ldots, A_n$ partition Ω, then

$$p_X(x) = \sum_{k=1}^{n} p_{X|A_k}(x)\,\mathbb{P}(A_k)$$

Given two rvs on same $\Omega, \mathbb{P}$ - say X and Y - define

$$p_{X|Y}(x|y) := \frac{\mathbb{P}(\{X=x\} \cap \{Y=y\})}{\mathbb{P}(\{Y=y\})} = \frac{p_{X,Y}(x,y)}{p_Y(y)}; \quad \forall x, y \text{ w/ } p_Y(y) > 0$$

For fixed y, $p_{X|Y}(x|y)$ defines a pmf "over x-values" - i.e.

$$p_{X|Y}(x|y) \geq 0 \ \forall x \quad \text{and} \quad \sum_{x \in X} p_{X|Y}(x|y) = 1$$

"$p_{X|Y}(x|y) =$ conditional pmf of X given $Y = y$"

Like the conditional "event-centered story", have a product rule of sorts

$$p_{X,Y}(x,y) = p_Y(y) p_{X|Y}(x|y) \quad \forall x \in X, y \in Y$$

OR

$$p_{X,Y}(x,y) = p_X(x) p_{Y|X}(y|x) \quad \forall x \in X, y \in Y$$

This expresses joint in terms of marginals + conditional(s).

Also have a total-probability rule of sorts:

$$p_X(x) = \sum_{y \in Y} p_Y(y) p_{X|Y}(x|y)$$

$$p_Y(y) = \sum_{x \in X} p_X(x) p_{Y|X}(y|x)$$

Moment Generating Function (MGF)

Given X, continuous or discrete, define MGF of X as

$$M_X(s) = \mathbb{E}(e^{sX}) \quad - \ s \text{ is a "variable"}$$

$M_X(s)$ is a function of s - need to be careful about its domain of definition

$$\mathbb{E}(X^k) = \frac{d^k}{ds^k} M_X(s) \Big|_{s=0}$$

Let

$$M_n = \frac{X_1 + \cdots + X_n}{n}, \quad X_k \text{ iid}$$

Saw

$$\mathbb{E}(M_n) = \mu \ \forall n; \quad Var(M_n) = \frac{\sigma^2}{n} \ \forall n$$

Chebyshev Inequality:

$$\mathbb{P}(\{|X - \mu| \geq c\}) \leq \frac{Var(X)}{c^2} \quad \forall c$$

From this, it follows that

$$\mathbb{P}(\{|M_n - \mu| \geq \varepsilon\}) \leq \frac{\sigma^2}{n\varepsilon^2} \quad \forall \varepsilon > 0$$

Consequence

$$\mathbb{P}(\{|M_n - \mu| \geq \varepsilon\}) \to 0 \quad \text{as} \quad n \to \infty \quad \forall \varepsilon > 0$$

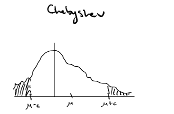
↑WLLN

Markov Inequality: If X is a nonnegative-valued rv, then for every $c > 0$,

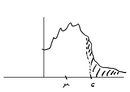
$$\mathbb{P}(\{X \geq c\}) \leq \frac{\mathbb{E}(X)}{c}$$

Think of these both as quantitative bounds on tail probabilities

Chebyshev



Markov



Another consequence of Chebyshev:

$$\mathbb{P}(\{|X - \mu| \geq k\sigma\}) \leq \frac{1}{k^2}$$ "probability of being k std deviation away from mean $\leq \frac{1}{k^2}$"

The kind of convergence taking place in WLLN **converges in probability**

Definition: Given a sequence of rv's Y_n, $1 \leq n < \infty$, and a number a, say Y_n converges in probability to a as $n \to \infty$ when $\forall \varepsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P}(\{|Y_n - a| \geq \varepsilon\}) = 0$$

WLLN just says "$M_n \to \mu$ in probability as $n \to \infty$"

Last thing: convergence w/ probability 1 of a sequence $Y_1, Y_2, \ldots$ of random variables.

Consider the sequence $\{Y_n : n > 0\}$,

Given some random variable Y,

$$\{\lim_{n \to \infty} Y_n = Y\}$$

is an event need to refer back to $\Omega, \mathbb{P}$, etc

Say $Y_n \to Y$ with probability 1 (w.p.1)

$Y_n \xrightarrow{\text{w.p.1}} Y$

$Y_n \xrightarrow{\text{a.s.}} Y$ a.s. = almost surely

when this event has probability 1.

Given $\Omega, \mathbb{P}$, and $X: \Omega \to \mathbb{R}$ a rv. Say X is a **continuous rv** when there exists a function $f_X(x)$ – called the probability density function (pdf) of X – such that "any" $\mathcal{V} \in \mathbb{R}$,

$$\mathbb{P}(\{X \in \mathcal{V}\}) = \int_{\mathcal{V}} f_X(x)\, dx$$ ← $f_X(x)$ has to be reasonable enough for integrals to make sense

- $f_X(x) \geq 0 \ \forall x$ (need this to ensure $\mathbb{P}(\{x \in \mathcal{V}\}) \geq 0 \ \forall \mathcal{V} \subset \mathbb{R}$)

- $\lim_{R\to\infty} \int_{-R}^{+R} f_X(x)\,dx = 1 \to \int_{-\infty}^{\infty} f_X(x)\,dx = \mathbb{P}(\{X \in (-\infty,\infty)\}) = 1$

- Given $x \in \mathbb{R}$, $f_X(x)$ is **NOT** $\mathbb{P}$(some event) – in particular,

$$f_X(x) \neq \mathbb{P}(\{X = x\})$$

Turns out $\mathbb{P}(\{X = x\}) = \mathbf{0} \ \forall x \in \mathbb{R}$ when X is a continuous random variable

Expected Value

The expected value of a continuous rv X w/ pdf $f_X(x)$:

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x f_X(x)\, dx$$ ← Caution: NOT always defined – integral might fail to exist

Expected Value Rule

Given X w/ pdf $f_X(x)$ and $Y = g(X)$, we have

$$\mathbb{E}[Y] = \int_{-\infty}^{+\infty} g(x) f_X(x)\, dx$$ ← enables $\mathbb{E}[Y]$ computation w/out finding $f_Y(y)$ or $p_Y(y)$

Variance

Variance of continuous rv:

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

By expected value rule, we have

$$\mathrm{Var}(X) = \int_{-\infty}^{+\infty} (X - \mathbb{E}[X])^2 f_X(x)\, dx$$

Also, as before,

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

Next, define – for **ANY** rv X (discrete OR continuous) the cumulative distribution function (cdf) by

$$F_X(x) = \mathbb{P}(\{X \leq x\}) \quad \forall x \in \mathbb{R}$$

Observation: If X is a continuous rv w/ pdf $f_X(x)$, then since

$$\mathbb{P}(\{X \leq x\}) = \int_{-\infty}^{x} f_X(t)\, dt$$

we have

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt \quad \text{and} \quad f_X(x) = \frac{d}{dx} F_X(x)$$

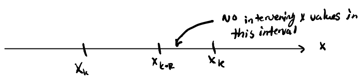
Discrete version: If X is a discrete rv w/ pmf $p_X(x)$ we have

$$F_X(x) = \sum_{\{x_k : x_k \leq x\}} p_X(x_k)$$ ← set of all possible X-values that don't exceed x

Can invert this formula to get $p_X(x)$ in terms of $F_X(x)$:

$$p_X(x_k) = F_X(x_k) - F_X(x_{k-1})$$

where x_{k-1} is the "**NEXT** largest value" of X below x_k.



General Properties of CDFs

① $\lim_{x\to 0} F_X(x) = 0$ and $\lim_{x\to\infty} F_X(x) = 1$

② When X is a continuous rv, $F_X(x)$ is continuous in x and differentiable "almost everywhere" (corners in $F_X(x)$ correspond to jumps in $f_X(x)$)

③ X is a discrete iff $F_X(x)$ is a piecewise constant.

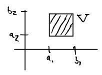
④ $F_X(x)$ is **monotonically** increasing in x.

$$x_1 \leq x_2 \Rightarrow F_X(x_1) \leq F_X(x_2)$$

Say X, Y rvs defined in same $\Omega, \mathbb{P}$ are jointly continuous w/ joint pdf $f_{X,Y}(x,y)$ when

$$\mathbb{P}(\{(X,Y) \in \mathcal{V}\}) = \iint_{\mathcal{V}} f_{X,Y}(x,y)\, dx\, dy \quad \forall \mathcal{V} \subset \mathbb{R}^2$$

Special case of a $\mathcal{V}: [a_1, b_1] \times [a_2, b_2]$



Then

$$\mathbb{P}(\{(X,Y) \in \mathcal{V}\}) = \int_{a_1}^{b_1} dx \int_{a_2}^{b_2} dy \left(f_{X,Y}(x,y)\right)$$

Again, have marginals

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\, dy \quad ; \quad f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\, dx$$

An official way to get this: Get $F_X(x)$ first then take $\frac{d}{dx} F_X(x)$

$$F_X(x) = \mathbb{P}(\{X \leq x\}) = \mathbb{P}(\{(X,Y) \in (-\infty, x] \times (-\infty, +\infty)\})$$

$$= \int_{-\infty}^{x} dt \int_{-\infty}^{+\infty} dy \left(f_{X,Y}(t,y)\right)$$

then

$$\frac{d}{dx} F_X(x) = \int_{-\infty}^{+\infty} dy \left(f_{X,Y}(x,y)\right)$$

Could also derive marginal formulas as follows:

$\forall \mathcal{V} \subset \mathbb{R}$, $\mathbb{P}(\{X \in \mathcal{V}\}) = \mathbb{P}(\{(X,Y) \in \mathcal{V} \times (-\infty,\infty)\})$

$$= \int_{\mathcal{V}} dx \int_{-\infty}^{+\infty} dy \left(f_{X,Y}(x,y)\right) = \int_{\mathcal{V}} \left(\int_{-\infty}^{+\infty} f_{X,Y}(x,y)\, dy\right) dx$$

Must be $f_X(x)$ integrate over $\mathcal{V}$ to get $\mathbb{P}(\{X \in V\})$

Other stuff

- $\int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy \left(f_{X,Y}(x,y)\right) = 1$
- Joint CDF; $F_{X,Y}(x,y) = \mathbb{P}(\{X \leq x\} \cap \{Y \leq y\}) = \int_{-\infty}^{x} ds \int_{-\infty}^{t} dt \left(f_{X,Y}(s,t)\right)$
- $f_{X,Y}(x,y) = \frac{\partial}{\partial x}\frac{\partial}{\partial y} F_{X,Y}(x,y)$

Conditional Stuff For Continuos Random Variables

Given a continuous rv X on $\Omega, \mathbb{P}$ and some event $A \subset \Omega$, the conditional pdf of X given A "defined" as follows:

For any $\mathcal{V} \subset \mathbb{R}$, we have

$$\mathbb{P}(\{X \in \mathcal{V}\} | A) = \int_{\mathcal{V}} f_{X|A}(x)\, dx$$

In general, no decent formula for $f_{X|A}(x)$ in terms of $f_X(x)$.
One way to compute it:

- First get conditional cdf of x given A

$$F_{X|A} = \mathbb{P}(\{X \leq x\} | A)$$

- Then take d/dx to get $f_X(x)$

However, if A is an event of the form $\{X \in \mathcal{W}\}$, and $\mathbb{P}(A) > 0$, we have

$$f_{X|A}(x) = \begin{cases} \dfrac{f_X(x)}{\mathbb{P}(\{X \in \mathcal{W}\})}, & \text{when } X \in \mathcal{W} \\ 0, & \text{otherwise} \end{cases}$$

How does this arise?

$$\mathbb{P}(\{X \in \mathcal{V}\} | A) = \frac{\mathbb{P}(\{X \in (\mathcal{V} \cap \mathcal{W})\})}{\mathbb{P}(\{X \in \mathcal{W}\})} = \frac{\int_{\mathcal{V}\cap\mathcal{W}} f_X(x)\, dx}{\mathbb{P}(\{X \in \mathcal{W}\})}$$

Total Probability Theorem in context of $f_{X|A}$:

If X is a continuous rv and $A_1, \ldots, A_n$ are events of positive probability that partition Ω, then

$$f_X(x) = \sum_{k=1}^{n} f_{X|A_k}(x)\, \mathbb{P}(A_k)$$

To see this: go via cdfs.

$$F_{X|A_k} = \frac{\mathbb{P}(\{X \leq x\} \cap A_k)}{\mathbb{P}(A_k)}$$

$$\frac{d}{dx} F_{X|A_k} = f_{X|A_k}(x)$$

By Total Probability Theorem,

$$F_X(x) = \mathbb{P}(\{X \leq x\}) = \sum_{k=1}^{n} F_{X|A_k} \mathbb{P}(A_k) \xrightarrow{\frac{d}{dx}} \sum_{k=1}^{n} f_{X|A_k} \mathbb{P}(A_k) = f_X(x)$$

Comment: this holds when A_k aren't of the special form $\{X \in W_k\}$!

Bottom line: conditional pdf of X given $Y = y$ is

$$f_{X|Y}(x,y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$ ← What you integrate over for any $x \in \mathcal{V}$ to get $\mathbb{P}(\{X \in \mathcal{V}\} | \{Y = y\})$

e.g.

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

Integrate over x or y to get

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X|Y}(x|y)\, dy \quad \text{OR} \quad f_Y(y) = \int_{-\infty}^{+\infty} f_{Y|X}(y|x) f_X(x)\, dx$$

- X is a continuous rv whose density $f_X(x)$ is concentrated on a single interval $a < x < b$. $a = -\infty$ and/or $b = +\infty$ allowed.
- $Y = g(X)$, g strictly monotonic and differentiable, implying that $f_Y(y)$ is concentrated on $(g(a), g(b))$ g increasing OR $(g(b), g(a))$ g decreasing
- Let h be the inverse function of g – defined only on $(g(a), g(b))$ or $(g(b), g(a))$ — h is also strictly monotonic and differentiable on its domain of definition

Then

$$f_Y(y) = \begin{cases} \left|\dfrac{dh(y)}{dy}\right| f_X(h(y)), & y \in (g(a), g(b)) \text{ OR } y \in (g(b), g(a)) \\ 0, & \text{else} \end{cases}$$

(increasing, decreasing)

Covariance

Given any X, Y rvs (discrete, continuous, whatever) defined on **same** probability space, the **covariance** of X and Y defined as

$$\mathrm{Cov}(X,Y) = \mathbb{E}\left((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

$$\mathrm{Cov}(X,X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \mathrm{Var}(X)$$

Terminology: When $\mathrm{Cov}(X,Y) = 0$; say X and Y are uncorrelated.

Fact: If X, Y independent, then X, Y uncorrelated

Terminology: Given X, Y, the correlation coefficient of X and Y is

$$\rho = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

Central Limit Theorem

Converging in the sense that for every z, $F_{Z_n}(z) \longleftrightarrow \Phi(z)$

Recall that if X_k iid w/ common μ, σ^2

$$M_n = \frac{X_1 + \cdots + X_n}{n} \Rightarrow \mathbb{E}(M_n) = \mu \text{ and } \mathrm{Var}(M_n) = \frac{\sigma^2}{n}$$

Form Z_n by renormalizing so $\mathbb{E}(Z_n) = 0$ and $\mathrm{Var}(Z_n) = 1 \ \forall n$.

$$Z_n = \frac{X + \cdots + X_n - n\mu}{\sqrt{n}\,\sigma} = \frac{\sqrt{n}}{\sigma}(M_n - \mu)$$

In this context, Z_n converges to, as $n \to \infty$, a Gaussian with mean $\mu = 0$ and variance $\sigma^2 = 1$.

Conditional Expectation Revisited

Terminology: $\mathbb{E}(X|Y)$ = conditional expectation of X given Y

Question: What is $\mathbb{E}(\mathbb{E}(X|Y))$?

Fact: Law of iterated expectations

$$\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|Y))$$

Idea: $\mathbb{E}(X|Y) = g(Y)$ for some function Y.

Thus

$$\mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}(g(Y))$$ ← use expected value rule to get this

Law of Iterated Expectations: $\mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}(X)$

For any function $h(Y)$,

- $\mathbb{E}(h(Y)|Y) = h(Y)$
- $\mathbb{E}(h(Y)X|Y) = h(Y)\mathbb{E}(X|Y)$

Can think of $\mathbb{E}(X|Y)$ as an estimator of X given Y.

In what sense does it "act like an estimator?"

- $\mathbb{E}(Y|Y) = \mathbb{E}(X)$ by law of iterated expectations
- The estimation error $X - \mathbb{E}(X|Y)$ is uncorrelated w/ the estimate $\mathbb{E}(X|Y)$ – in fact, $X - \mathbb{E}(X|Y)$ is uncorrelated with Y – More generally, w/ any function $h(Y)$

Fact (Major): $\mathbb{E}(X|Y)$ is the function of Y that minimizes $\mathbb{E}((X - h(Y))^2)$

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x) f_X(x)}{f_Y(y)} = \frac{f_{Y|X}(y|x) f_X(x)}{\int_{-\infty}^{+\infty} f_{Y|X}(y|x) f_X(x)\, dx}$$ } Continuous Baye's Rule