Probability review

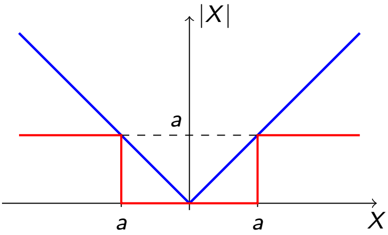
Alejandro Ribeiro
Dept. of Electrical and Systems Engineering
University of Pennsylvania
aribeiro@seas.upenn.edu
http://www.seas.upenn.edu/users/~aribeiro/

September 22, 2010

Markov's inequality

- RV X with finite expected value $\mathbb{E}(X) < \infty$
- Markov's inequality states $\Rightarrow \mathrm{P}\left[|X| \geq a\right] \leq \dfrac{\mathbb{E}(|X|)}{a}$
- $\mathbb{I}\left\{|X| \geq a\right\} = 1$ when $X \geq a$ and 0 else. Then (figure to the right)

$$a\mathbb{I}\left\{|X| \geq a\right\} \leq |X|$$



- Expected value. Linearity of $\mathbb{E}\left[\cdot\right]$

$$a\mathbb{E}(\mathbb{I}\left\{|X| \geq a\right\}) \leq \mathbb{E}(|X|)$$

- Indicator function's expectation = Probability of event

$$a\mathrm{P}\left[|X| \geq a\right] \leq \mathbb{E}(|X|)$$

Chebyshev's inequality

- RV X with finite mean $\mathbb{E}(X) = \mu$ and variance $\mathbb{E}\left[(X-\mu)^2\right] = \sigma^2$
- Chebyshev's inequality $\Rightarrow \mathrm{P}\left[|X-\mu| \geq k\right] \leq \dfrac{\sigma^2}{k^2}$
- Markov's inequality for the RV $Z = (X-\mu)^2$ and constant $a = k^2$

$$\mathrm{P}\left[(X-\mu)^2 \geq k^2\right] = \mathrm{P}\left[|Z| \geq k^2\right] \leq \frac{\mathbb{E}\left[|Z|\right]}{k^2} = \frac{\mathbb{E}\left[(X-\mu)^2\right]}{k^2}$$

- Notice that $(X-\mu)^2 \geq k^2$ if and only if $|X-\mu| \geq k$ thus

$$\mathrm{P}\left[|X-\mu| \geq k\right] \leq \frac{\mathbb{E}\left[(X-\mu)^2\right]}{k^2}$$

- Chebyshev's inequality follows from definition of variance

Comments & observations

- Markov and Chebyshev's inequalities hold for all RVs
- If absolute expected value is finite $\mathbb{E}\left[|X|\right] < \infty$

  $\Rightarrow$ RV's cdf decreases at least linearly (Markov's)
- If mean $\mathbb{E}(X)$ and variance $\mathbb{E}\left[(X-\mu)^2\right]$ are finite

  $\Rightarrow$ RV's cdf decreases at least quadratically (Chebyshev's)
- Most cdfs decrease exponentially (e.g. e^{-x^2} for normal)

  $\Rightarrow$ linear and quadratic bounds are loose but still useful

Limits in probability

Markov and Chebyshev's Inequalities

Limits in probability

Limit theorems

Conditional probabilities

Conditional expectation

Limits

- Sequence of RVs $X_{\mathbb{N}} = X_1, X_2, \ldots, X_n, \ldots$
- Distinguish between stochastic process $X_{\mathbb{N}}$ and realizations $x_{\mathbb{N}}$
- Say something about X_n for n large? $\Rightarrow$ Not clear, X_n is a RV
- Say something about x_n for n large? $\Rightarrow$ Certainly, look at $\lim\limits_{n\to\infty} x_n$
- Say something about $\mathrm{P}[X_n]$ for n large? $\Rightarrow$ Yes, $\lim\limits_{n\to\infty} \mathrm{P}[X_n]$
- Translate what we now about regular limits to definitions for RVs
- Can start from convergence of sequences: $\lim_{n\to\infty} x_n$
  - Sure and almost sure convergence
- Or from convergence of probabilities: $\lim_{n\to\infty} \mathrm{P}[X_n]$
  - Convergence in probability, mean square sense and distribution

Convergence of sequences and sure convergence

- Denote sequence of variables $x_{\mathbb{N}} = x_1, x_2, \ldots, x_n, \ldots$
- Sequence $x_{\mathbb{N}}$ converges to the value x if given any $\epsilon > 0$

  $\Rightarrow$ There exists n_0 such that for all $n > n_0$, $|x_n - x| < \epsilon$
- Sequence x_n comes close to its limit $\Rightarrow |x_n - x| < \epsilon$
- And stays close to its limit $\Rightarrow$ for all $n > n_0$

- Stochastic process (sequence of RVs) $X_{\mathbb{N}} = X_1, X_2, \ldots, X_n, \ldots$
- Realizations of $X_{\mathbb{N}}$ are sequences $x_{\mathbb{N}}$
- We say SP $X_{\mathbb{N}}$ converges surely to RV X if $\Rightarrow \lim_{n\to\infty} x_n = x$
- For all realizations $x_{\mathbb{N}}$ of $X_{\mathbb{N}}$

- Not really adequate. Even an event that happens with vanishingly small probability prevents sure convergence

Almost sure convergence

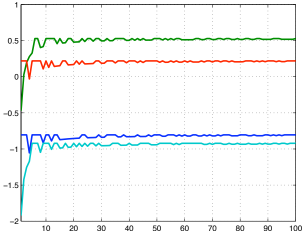
- RV X and stochastic process $X_{\mathbb{N}} = X_1, X_2, \ldots, X_n, \ldots$
- We say SP $X_{\mathbb{N}}$ converges almost surely to RV X if

$$\mathrm{P}\left[\lim_{n\to\infty} X_n = X\right] = 1$$

- Almost all sequences converge, except for a set of measure 0
- Almost sure convergence denoted as $\Rightarrow \lim_{n\to\infty} X_n = X \quad \text{a.s.}$
- Limit X is a random variable

Example

- $X_0 \sim \mathcal{N}(0,1)$ (normal, mean 0, variance 1)
- Z_n Bernoulli parameter p
- Define $\Rightarrow X_n = X_0 - \dfrac{Z_n}{n}$
- $Z_n/n \to 0$, then $\lim_{n\to\infty} X_n = X_0$ a.s.



Convergence in probability

- We say SP $X_{\mathbb{N}}$ converges in probability to RV X if for any $\epsilon > 0$

$$\lim_{n\to\infty} \mathrm{P}\left[|X_n - X| < \epsilon\right] = 1$$

- Probability of distance $|X_n - X|$ becoming smaller than ϵ tends to 1
- Statement is about probabilities, not about processes
- The probability converges
- Realizations $x_{\mathbb{N}}$ of $X_{\mathbb{N}}$ might or might not converge

- Limit and probability interchanged with respect to a.s. convergence

- a.s. convergence implies convergence in probability
  - If $\lim_{n\to\infty} X_n = X$ then for any $\epsilon > 0$ there is n_0 such that $|X_n - X| < \epsilon$ for all $n \geq n_0$
  - This is true for all almost all sequences then $\mathrm{P}\left[|X_n - X| < \epsilon\right] \to 1$

Convergence in probability (continued)

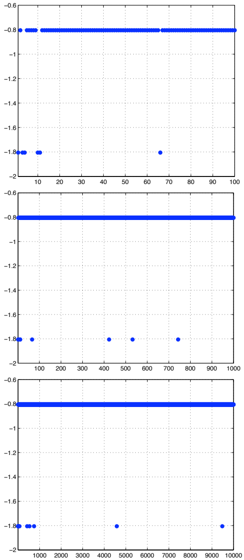
Example

- $X_0 \sim \mathcal{N}(0,1)$ (normal, mean 0, variance 1)
- Z_n Bernoulli parameter $1/n$
- Define $\Rightarrow X_n = X_0 - Z_n$

- X_n converges in probability to X_0 because

$$\begin{aligned} \mathrm{P}\left[|X_n - X_0| < \epsilon\right] &= \mathrm{P}\left[|Z_n| < \epsilon\right] \\ &= 1 - \mathrm{P}\left[Z_n = 1\right] \\ &= 1 - \frac{1}{n} \to 1 \end{aligned}$$

- Plot of path x_n up to $n = 10^2$, $n = 10^3$, $n = 10^4$
- $Z_n = 1$ becomes ever rarer but still happens



Difference between a.s. and p

- Almost sure convergence implies that almost all sequences converge
- Convergence in probability does not imply convergence of sequences

- Latter example: $X_n = X_0 - Z_n$, Z_n is Bernoulli with parameter $1/n$
- As we've seen it converges in probability

$$\mathrm{P}\left[|X_n - X_0| < \epsilon\right] = 1 - \frac{1}{n} \to 1$$

- But for almost all sequences, the $\lim_{n\to\infty} X_n$ does not exist

- Almost sure convergence $\Rightarrow$ disturbances stop happening
- Convergence in prob. $\Rightarrow$ disturbances happen with vanishing freq.
- Difference not irrelevant.
  - Interpret Y_n as rate of change in savings
  - with a.s. convergence risk is eliminated
  - with convergence in probability risk decreases but does not disappear

Mean square convergence

- We say SP $X_{\mathbb{N}}$ converges in mean square to RV X if

$$\lim_{n\to\infty} \mathbb{E}\left[|X_n - X|^2\right] = 0$$

- Sometimes (very) easy to check
- Convergence in mean square implies convergence in probability
- From Markov's inequality

$$\mathrm{P}\left[|X_n - X| \geq \epsilon\right] = \mathrm{P}\left[|X_n - X|^2 \geq \epsilon^2\right] \leq \frac{\mathbb{E}\left[|X_n - X|^2\right]}{\epsilon^2}$$

- If $X_n \to X$ in mean square sense, $\mathbb{E}\left[|X_n - X|^2\right]/\epsilon^2 \to 0$ for all ϵ
- Almost sure and mean square $\not\Rightarrow$ neither implies the other

Convergence in distribution

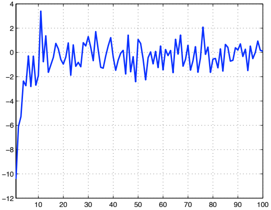
- Stochastic process $X_{\mathbb{N}}$. Cdf of X_n is $F_n(x)$
- The SP converges in distribution to RV X with distribution $F_X(x)$ if

$$\lim_{n\to\infty} F_n(x) = F_X(x)$$

- For all x at which $F_X(x)$ is continuous
- Again, no claim about individual sequences, just the cdf of X_n
- Weakest form of convergence covered,
- Implied by almost sure, in probability, and mean square convergence

Example

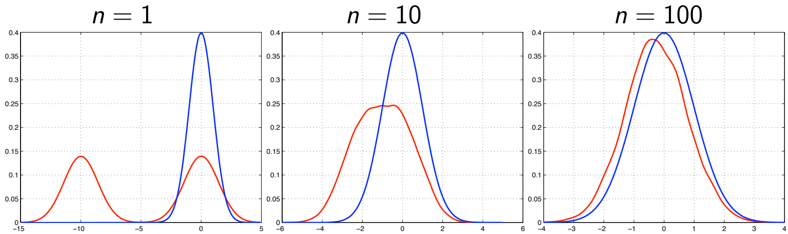
- $Y_n \sim \mathcal{N}(0,1)$
- Z_n Bernoulli parameter p
- Define $\Rightarrow X_n = Y_n - 10Z_n/n$
- $Z_n/n \to 0$, then $\lim_{n\to\infty} F_n(x) = \mathcal{N}(0,1)$



Convergence in distribution (continued)

- Individual sequences x_n do not converge in any sense

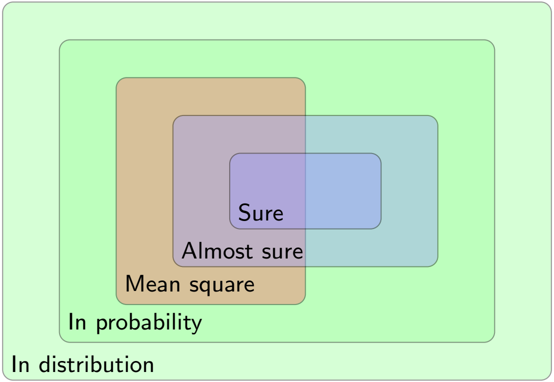
  $\Rightarrow$ It is the distribution that converges



- As the effect of Z_n/n vanishes pdf of X_n converges to pdf of Y_n
  - Standard normal $\mathcal{N}(0, 1)$

Implications

- Sure $\Rightarrow$ almost sure $\Rightarrow$ in probability $\Rightarrow$ in distribution
- Mean square $\Rightarrow$ in probability $\Rightarrow$ in distribution
- In probability $\Rightarrow$ in distribution



Sum of independent identically distributed RVs

- Independent identically distributed (i.i.d.) RVs $X_1, X_2, \ldots, X_n, \ldots$
- Mean $\mathbb{E}[X_n] = \mu$ and variance $\mathbb{E}\left[(X_n - \mu)^2\right] = \sigma^2$ for all n
- What happens with sum $S_N := \sum_{n=1}^{N} X_n$ as N grows?

- Expected value of sum is $\mathbb{E}[S_N] = N\mu \;\Rightarrow$ Diverges if $\mu \neq 0$
- Variance is $\mathbb{E}\left[(S_N - N\mu)^2\right] = N\sigma$

  $\Rightarrow$ Diverges if $\sigma \neq 0$ (alwyas true unless X_n is a constant)

- One interesting normalization $\Rightarrow \bar{X}_N := (1/N)\sum_{n=1}^{N} x_n$
- Now $\mathbb{E}[Z_N] = \mu$ and $\text{var}[Z_N] = \sigma^2/N$
- Law of large numbers (weak and strong)

- Another interesting normalization $\Rightarrow Z_N := \dfrac{\sum_{n=1}^{N} x_n - N\mu}{\sigma\sqrt{N}}$
- Now $\mathbb{E}[Z_N] = 0$ and $\text{var}[Z_N] = 1$ for all values of N
- Central limit theorem

Weak law of large numbers

- i.i.d. sequence or RVs $X_1, X_2, \ldots, X_n, \ldots$ with mean $\mu = \mathbb{E}[X_n]$
- Define sample average $\bar{X}_N := (1/N)\sum_{n=1}^{N} x_n$
- Weak law of large numbers
- Sample average $\bar{X}_N$ converges in probability to $\mu = \mathbb{E}[X_n]$

$$\lim_{N\to\infty} \mathrm{P}\left[|\bar{X}_N - \mu| > \epsilon\right] = 1, \quad \text{for all } \epsilon > 0$$

- Strong law of large numbers
- Sample average $\bar{X}_N$ converges almost surely to $\mu = \mathbb{E}[X_n]$

$$\mathrm{P}\left[\lim_{N\to\infty} \bar{X}_N = \mu\right] = 1$$

- Strong law implies weak law. Can forget weak law if so wished

Proof of weak law of large numbers

- Weak law of large numbers is very simple to prove

Proof.

- Variance of $\bar{X}_n$ vanishes for N large

$$\text{var}\left[\bar{X}_N\right] = \frac{1}{N^2}\sum_{n=1}^{n} \text{var}\left[X_n\right] = \frac{\sigma^2}{N} \to 0$$

- But, what is the variance of $\bar{X}_N$?

$$0 \leftarrow \frac{\sigma^2}{N} = \text{var}\left[\bar{X}_N\right] = \mathbb{E}\left[(\bar{X}_n - \mu)^2\right]$$

- Then, $|\bar{X}_N - \mu|$ converges in mean square sense

  $\Rightarrow$ Which implies convergence in probability $\square$

- Strong law is a little more challenging

Central limit theorem (CLT)

Theorem

- *i.i.d. sequence of RVs* $X_1, X_2, \ldots, X_n, \ldots$
- *Mean* $\mathbb{E}\left[X_n\right] = \mu$ *and variance* $\mathbb{E}\left[(X_n - \mu)^2\right] = \sigma^2$ *for all* n
- *Then* $\Rightarrow \lim_{N\to\infty} \mathrm{P}\left[\frac{\sum_{n=1}^{N} x_n - N\mu}{\sigma\sqrt{N}} \leq x\right] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-u^2/2}\, du$

- Former statement implies that for N sufficiently large

$$Z_N := \frac{\sum_{n=1}^{N} x_n - N\mu}{\sigma\sqrt{N}} \sim \mathcal{N}(0, 1)$$

- $\sim$ means "distributed like"
- Z_N converges in distribution to a standard normal RV

CLT (continued)

- Equivalently can say $\Rightarrow \sum_{n=1}^{N} x_n \sim \mathcal{N}(N\mu, N\sigma^2)$
- Sum of large number of i.i.d. RVs has a normal distribution
  - Cannot take a meaningful limit here.
  - But intuitively, this is what the CLT states

Example

- Binomial RV X with parameters (n, p)
- Write as $X = \sum_{i=1}^{n} X_i$ with X_i Bernoulli with parameter p
- Mean $\mathbb{E}[X_i] = p$ and variance $\text{var}[X_i] = p(1-p)$
- For sufficiently large $n \Rightarrow X \sim \mathcal{N}(n\mu, np(1-p))$

Conditional probabilities

Conditional pmf and cdf for discrete RVs

- Recall definition of conditional probability for events E and F

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)}$$

- Change in likelihoods when information is given, renormalization
- Define the conditional pmf of RV X given Y as (both RVs discrete)

$$p_{X|Y}(x \mid y) = \mathrm{P}\left[X = x \mid Y = y\right] = \frac{\mathrm{P}[X = x, Y = y]}{\mathrm{P}[Y = y]}$$

- Which we can rewrite as

$$p_{X|Y}(x \mid y) = \frac{\mathrm{P}[X = x, Y = y]}{\mathrm{P}[Y = y]} = \frac{p_{XY}(x, y)}{p_Y(y)}$$

- Pmf for random variable x, given parameter y ("Y not random anymore")
- Define conditional cdf as (a range of X conditional on a value of Y)

$$F_{X|Y}(x \mid y) = \mathrm{P}\left[X \leq x \mid Y = y\right] = \sum_{z \leq x} p_{X|Y}(z \mid y)$$

Example

Example

- Independent Bernoulli Y and Z, variable $X = Y + Z$
- Conditional pmf of X given Y? For $X = 0$, $Y = 0$

$$p_{X|Y}(X = 0 \,\big|\, Y = 0) = \frac{\mathrm{P}\left[X = 0, Y = 0\right]}{\mathrm{P}\left[Y = 0\right]} = \frac{(1-p)^2}{1-p} = 1 - p$$

- Or, from joint and marginal pdfs (just a matter of definition)

$$p_{X|Y}(X = 0 \,\big|\, Y = 0) = \frac{p_{XY}(0,0)}{p_Y(0)} = \frac{(1-p)^2}{1-p} = 1 - p$$

- Can compute the rest analogously

$$\begin{aligned} &p_{X|Y}(0|0) = (1-p), \quad &&p_{X|Y}(1|0) = p, \quad &&p_{X|Y}(2|0) = 0 \\ &p_{X|Y}(0|1) = 0, \quad &&p_{X|Y}(1|1) = 1-p, \quad &&p_{X|Y}(2|1) = p \end{aligned}$$

Conditional pdf and cdf for continuous RVs

- Define conditional pdf of RV X given Y as (both RVs continuous)

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

- For motivation, define intervals $\Delta x = [x, x{+}dx]$ and $\Delta y = [y, y{+}dy]$
- Can approximate conditional probability $\mathrm{P}\left[X \in \Delta x \,\middle|\, Y \in \Delta y\right]$ as

$$\mathrm{P}\left[X \in \Delta x \,\middle|\, Y \in \Delta y\right] = \frac{\mathrm{P}\left[X \in \Delta x, Y \in \Delta y\right]}{\mathrm{P}\left[Y \in \Delta y\right]} \approx \frac{f_{XY}(x, y)dxdy}{f_Y(y)dy}$$

- From definition of conditional pdf it follows after simplifying terms

$$\mathrm{P}\left[X \in \Delta x \,\middle|\, Y \in \Delta y\right] \approx f_{X|Y}(x \,|\, y)dx$$

- Which is what we would expect of a density
- Conditional cdf defined as $\Rightarrow F_{X|Y}(x) = \int_{-\infty}^{x} f_{X|Y}(u \,|\, y)du$

Example: Communications channel

- Random message (RV) Y, transmit signal y (realization of Y)
- Received signal is $x = y + z$ (z realization of random noise)
- Can model communication system as a relation between RVs

$$X = Y + Z$$

- Model communication noise as $Z \sim \mathcal{N}(0, \sigma^2)$ independent of Y
- Conditional pdf of X given Y. Use definition:

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

- Problem is we don't know $f_{XY}(x, y)$. Have to calculate
- Computing conditional probs. typically easier than computing joints

Example: Communications channel (continued)

- If $Y = y$ is given, then "Y not random anymore" (Dorothy's principle)

  $\Rightarrow$ It still is random in reality, we are thinking of it as given
- If Y were not random, say $Y = y$ with y given then ...

$$X = y + Z$$

- Cdf of X, now easily obtained

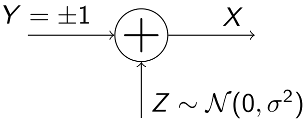
$$\mathrm{P}\left[X \leq x\right] = \mathrm{P}\left[y + Z \leq x\right] = \mathrm{P}\left[Z \leq x - y\right] = \int_{-\infty}^{x-y} p_Z(z)\, dz$$

- But since Z is normal with 0 mean and variance σ^2

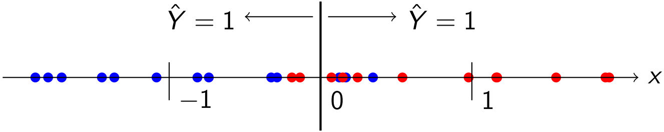
$$\mathrm{P}\left[X \leq x\right] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x-y} e^{-z^2/2\sigma^2}\, dz = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-(z-y)^2/2\sigma^2}\, dz$$

- X is normal with mean y and variance σ^2

- Conditioning is a common tool to compute probabilities
- Message 1 (prob. p) $\Rightarrow$ Transmit $Y = 1$
- Message 2 (prob. q) $\Rightarrow$ Transmit $Y = -1$
- Received signal $\Rightarrow X = Y + Z$



- Decoding rule $\Rightarrow \hat{Y} = 1$ if $X \geq 0$, $\hat{Y} = 0$ if $X < 0$
- What is the probability of error, $P_e := \mathrm{P}\left[\hat{Y} \neq Y\right]$?
- Red dots to the left and blue dots to the right are errors



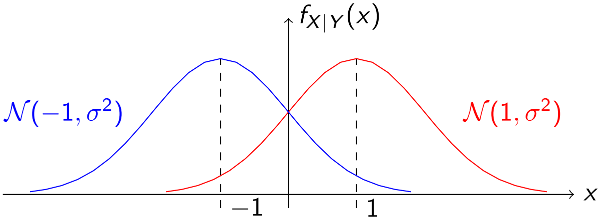
Output pdf

- From communications channel example we know
- If $Y = 1$, then $X \sim \mathcal{N}(1, \sigma^2)$, conditional pdf is

$$f_{X|Y}(x, 1) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-1)^2/2\sigma^2}$$

- If $Y = -1$, then $X \sim \mathcal{N}(-1, \sigma^2)$, conditional pdf is

$$f_{X|Y}(x, -1) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x+1)^2/2\sigma^2}$$



Error probability

- Write probability of error by conditioning on $Y = \pm 1$ (total probability)

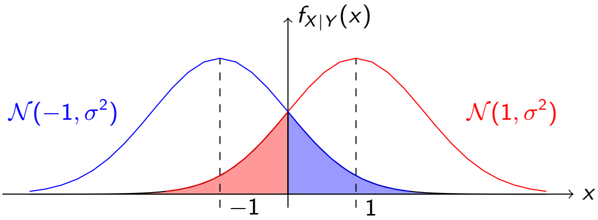
$$P_e = \mathrm{P}\{\hat{Y} \neq Y \,|\, Y = 1\}\mathrm{P}\{Y = 1\} + \mathrm{P}\{\hat{Y} \neq Y \,|\, Y = -1\}\mathrm{P}\{Y = -1\}$$
$$= \mathrm{P}\{\hat{Y} = -1 \,|\, Y = 1\}\, p \qquad + \mathrm{P}\{\hat{Y} = 1 \,|\, Y = -1\}\, q$$

- But according to the decision rule

$$P_e = \mathrm{P}\{X < 0 \,|\, Y = 1\}p + \mathrm{P}\{X \geq 0 \,|\, Y = -1\}q$$

- But X given Y is normally distributed, then

$$P_e = \frac{p}{\sqrt{2\pi}\sigma}\int_0^{\infty} e^{-(x-1)^2/2\sigma^2} + \frac{q}{\sqrt{2\pi}\sigma}\int_{-\infty}^{0} e^{-(x+1)^2/2\sigma^2} = \frac{q}{\sqrt{2\pi}\sigma}\int_{-\infty}^{0} e^{-x^2/2\sigma^2}$$



Definition of conditional expectation

- For continuous RVs X, Y define conditional expectation as

$$\mathbb{E}\left[X \mid y\right] = \int_{-\infty}^{\infty} x\, f_{X|Y}(x|y)\, dx$$

  For discrete RVs X, Y conditional expectation is

$$\mathbb{E}\left[X \mid y\right] = \sum_{x} x\, p_{X|Y}(x|y)$$

- Defined for given y $\Rightarrow \mathbb{E}\left[X \mid y\right]$ is a value
- All possible values y of Y $\Rightarrow$ random variable $\mathbb{E}\left[X \mid Y\right]$
- Y is RV, $\mathbb{E}\left[X \mid y\right]$ value associated with outcome $Y = y$

Double expectation

- If $\mathbb{E}\left[X \mid Y\right]$ is a RV, can compute expected value $\mathbb{E}_Y\left[\mathbb{E}_X\left(X \mid Y\right)\right]$
  - Subindices are for clarity purposes, innermost expectation is with respect to X, outermost with respect to Y
- What is $\mathbb{E}_Y\left[\mathbb{E}_X\left(X \mid Y\right)\right]$? Not surprisingly $\Rightarrow \mathbb{E}\left[X\right] = \mathbb{E}_Y\left[\mathbb{E}_X\left(X \mid Y\right)\right]$
- Show for discrete RVs (write integrals for continuous)

$$\mathbb{E}_Y\left[\mathbb{E}_X\left(X \mid Y\right)\right] = \sum_y \mathbb{E}_X\left(X \mid y\right) p_Y(y) = \sum_y \left[\sum_x x\, p_{X|Y}(x|y)\right] p_Y(y)$$
$$= \sum_x x \left[\sum_y p_{X|Y}(x|y) p_Y(y)\right] = \sum_x x \left[\sum_y p_{X,Y}(x,y)\right]$$
$$= \sum_x x p_X(x) = \mathbb{E}\left[X\right]$$

- Yields a method to compute expected values

$\Rightarrow$ Condition on $Y = y$ $\Rightarrow X \mid y$
$\Rightarrow$ Compute expected value over X for given y $\Rightarrow \mathbb{E}_X\left(X \mid y\right)$
$\Rightarrow$ Compute expected value over all values y of Y $\Rightarrow \mathbb{E}_Y\left[\mathbb{E}_X\left(X \mid Y\right)\right]$

Example

- Seniors get $A = 4$ with prob. 0.5, $B = 3$ with prob. 0.5
- Juniors get $B = 3$ with prob. 0.6, $B = 2$ with prob. 0.4
- Exchange student's standing: senior (junior) with prob. 0.7 (0.3)
- Expectation of X = exchange student's grade?

- Begin conditioning on standing

$$\mathbb{E}\left[X \,\middle|\, \text{Senior}\right] = 0.5 \times 4 + 0.5 \times 3 = 3.5$$
$$\mathbb{E}\left[X \,\middle|\, \text{Junior}\right] = 0.6 \times 3 + 0.4 \times 2 = 2.6$$

- Now sum over standing's probability

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}\left[X \,\middle|\, \text{Senior}\right] \mathrm{P}[\text{Senior}] + \mathbb{E}\left[X \,\middle|\, \text{Junior}\right] \mathrm{P}[\text{Junior}] \\ &= 3.5 \times 0.7 \qquad\qquad\qquad\qquad + 2.6 \times 0.3 \\ &= 3.23\end{aligned}$$

Conditioning to compute expectations

- As with probabilities conditioning is useful to compute expectations.
  - $\Rightarrow$ Spreads difficulty into simpler problems

Example

- A baseball player hits X_i runs per game
- Expected number of runs is $\mathbb{E}[X_i] = \mathbb{E}[X]$ independently of game
- Player plays N games in the season. N is random (playoffs, injuries?)
- Expected value of number of games is $\mathbb{E}[N]$
- What is the expected number of runs in the season ? $\Rightarrow \mathbb{E}\left[\sum_{i=1}^{N} X_i\right]$
- Both, N and X_i are random

Sum of random number of random quantities

Step 1: Condition on $N = n$ then

$$\left[\left(\sum_{i=1}^{N} X_i\right) \,\middle|\, N = n\right] = \sum_{i=1}^{n} X_i$$

Step 2: Compute expected value with respect to X_i

$$\mathbb{E}_{X_i}\left[\left(\sum_{i=1}^{N} X_i\right) \,\middle|\, N = n\right] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = n\mathbb{E}\left[X\right]$$

Second equality possible because n is a number (not a RV like N)

Step 3: Conpute expected value with respect to values n of N

$$\mathbb{E}_N\left[\mathbb{E}_{X_i}\left[\left(\sum_{i=1}^{N} X_i\right) \,\middle|\, N\right]\right] = \mathbb{E}_N\left[N\mathbb{E}\left[X\right]\right] = \mathbb{E}\left[N\right]\mathbb{E}\left[X\right]$$

Yielding result $\Rightarrow \mathbb{E}\left[\sum_{i=1}^{N} X_i\right] = \mathbb{E}\left[N\right]\mathbb{E}\left[X\right]$