Probability space $(\Omega, \{\text{events}\}, P)$

Expectation operator $E : \{\text{rv's}\} \to \mathbb{R}$

Examples of distributions on Ω discrete, (absolutely) cts

$$E[g(\omega)] = \sum_{\omega} g(\omega) f(\omega) \quad ; \quad E[g(\omega)] = \int_{\Omega} g(\omega) f(\omega)\, d\omega$$

rv, pf, pdf

discrete
{Bernoulli(p), binomial(n,p), geometric(p), negative binomial,
Poisson(λ), hypergeometric, etc...

{normal, gamma, chi-squared, Student-t, F, exponential
cts

The uniform can be discrete or cts.

$Y : \Omega \to \mathbb{R}$ generates a probability space
$(\mathbb{R}, \{\text{subsets of } \mathbb{R}\}, P_Y)$

via $P_Y(B) = P(\{Y \in B\})$, where $\{Y \in B\} = \{\omega : Y(\omega) \in B\}$.

Here B is a (Borel) subset of $\mathbb{R}$.

There are discrete & cts rv's and other types.

Random vectors also generate probability spaces, but in higher dimensions ($\mathbb{R}^k$, $k > 1$).

Basic inequalities

Triangle inequality: $E(|X+Y|) \leq E(|X|) + E(|Y|)$

Markov's inequality: $P(|X| \geq c) \leq \frac{E(|X|)}{c}$

Jensen's inequality: g convex $\Rightarrow E[g(X)] \geq g[E(X)]$

Proof of Jensen $\quad g$ convex $\Rightarrow \forall\ x_0 \quad \exists c = c(x_0) \ni$

$$g(x) \geq g(x_0) + c(x - x_0)$$

Take $x_0 = E(X)$. We then have

$$g(X) \geq g(E(X)) + c(X - E(X))$$

$$\Rightarrow E[g(X)] \geq g[E(X)] + c[E(X) - E(X)] = g[E(X)]$$

qed

Boole's Inequality $\quad P(\bigcup_k A_k) \leq \sum_k P(A_k)$

Proposition $\quad Y \geq 0$ & $E(Y) = 0 \Rightarrow P(Y=0) = 1 \quad$ (ie $Y \overset{as}{=} 0$)

The proof uses Boole & Markov.

r vec

mean $\quad \underset{\sim}{\mu} = E(\underset{\sim}{X}) = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_k) \end{pmatrix}$; $\quad \Sigma = Var(X) = \{cov(X_i, X_j)\}_{i,j=1}^{k}$

Lemma (i) $E(A\underset{\sim}{X} + \underset{\sim}{b}) = A\,E(\underset{\sim}{X}) + \underset{\sim}{b}$

(ii) $\mathrm{Var}(A\underset{\sim}{X} + \underset{\sim}{b}) = A\,\Sigma\,A'$

Consequence $\mathrm{Var}\left(\sum_{i=1}^{k} X_i\right) = \sum_{i,j} \mathrm{cov}(X_i, X_j)$

Consequence $\underset{\sim}{c}'\,\Sigma\,\underset{\sim}{c} \geq 0, \quad \forall\, \underset{\sim}{c}$

Assume $\underset{\sim}{c}'\,\Sigma\,\underset{\sim}{c} > 0, \quad \forall\, \underset{\sim}{c} \neq \underset{\sim}{0}$

Σ is <u>positive definite</u> $(\Sigma > 0)$

$$\left.\begin{array}{l} -\ \Sigma = \Sigma' \\ \underset{\sim}{c}'\Sigma\underset{\sim}{c} > 0, \ \forall\, \underset{\sim}{c} \neq \underset{\sim}{0} \end{array}\right\} \Rightarrow \begin{array}{l} \Sigma = T\,T' \\ = Q\,D\,Q' \end{array}$$

triangular

diagonal (>0 elements)

$\perp$

Linear prediction

$\underset{\sim}{X}$ ← predictor variables

Y ← rv to predict

$\hat{Y}$ ← predictor = f'n of $\underset{\sim}{X}$

Assume 0 means & only consider linear predictors

$$\hat{Y} = \underset{\sim}{a}'\,\underset{\sim}{X}$$

Set $\Sigma_{\underset{\sim}{X}\underset{\sim}{X}} = E(\underset{\sim}{X}\underset{\sim}{X}')$ $(= Var(\underset{\sim}{X})$ $\because$ 0 means$)$

product moment matrix

$\Sigma_{Y\underset{\sim}{X}} = E(Y\underset{\sim}{X})$

Choose $\underset{\sim}{a}$ to minimize the mean squared error

$$MSE(\underset{\sim}{a}) = E(Y - \underset{\sim}{a}'\underset{\sim}{X})^2$$

$$= E(Y^2) + \underset{\sim}{a}'\Sigma_{\underset{\sim}{X}\underset{\sim}{X}}\underset{\sim}{a} - 2\underset{\sim}{a}'\Sigma_{Y\underset{\sim}{X}}$$

$$\frac{\partial MSE(\underset{\sim}{a})}{\partial \underset{\sim}{a}'} = \underset{\sim}{0}'$$

yields the equation

$$\Sigma_{\underset{\sim}{X}\underset{\sim}{X}}\underset{\sim}{a} = \Sigma_{Y\underset{\sim}{X}} \qquad (*)$$

Any solution of (*) yields a best linear predictor (see Theorem 2.8.1). If $\Sigma_{\underset{\sim}{X}\underset{\sim}{X}}^{-1}$ exists then there is only one solution of (*).

Remark The case of non-zero means yields $\hat{Y} = E(Y) + \underset{\sim}{\hat{a}}'(\underset{\sim}{X} - \underset{\sim}{\mu})$

sol'n of (*)

$E(\underset{\sim}{X})$

The axioms for E include an important statement which is normally a theorem when starting a probability course from Kolmogorov's axioms for P.

<u>Lebesgue Monotone Convergence Theorem (MCT)</u>

$$0 \le X_n \uparrow X \implies E(X_n) \to E(X)$$

Without too much effort we may use this to prove (we will provide a proof later):

<u>Lebesgue Dominated Convergence Theorem (DCT)</u>

$X_n \to X$ + $|X_n| \le W$, $\forall n$ with $E(W) < \infty$

then $\underline{E(X_n) \to E(X)}$.

<u>Remark</u> In the DCT we may in fact show $E(|X_n - X|) \to 0$, which is called convergence in $\mathcal{L}_1$. A notation for this is $X_n \xrightarrow{L_1} X$

<u>Try these 2</u> – $X_n \xrightarrow{ms} X \implies X_n \xrightarrow{L_1} X \implies E(X_n) \to E(X)$

– derive Chebyshev from Markov