

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

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

Examples of distributions on  $\Omega$  discrete, (absolutely) cts

$$E[g(\omega)] = \sum_{\omega} g(\omega) \underbrace{f(\omega)}_{\substack{\uparrow \\ \text{PT}}} ; E[g(\omega)] = \int_{\Omega} g(\omega) \underbrace{f(\omega)}_{\substack{\uparrow \\ \text{pdf}}} d\omega$$

discrete

rv

PT

{Bernoulli(p), binomial(n,p), geometric(p), negative binomial,  
Poisson( $\lambda$ ), hypergeometric, etc...}

{normal, gamma, chi-squared, Student-t, F, exponential  
etc}

The uniform can be discrete or cts.

$Y: \Omega \rightarrow \mathbb{R}$  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  $g$  convex  $\Rightarrow \forall x_0 \exists c = c(x_0) \Rightarrow$

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

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

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

$$\Rightarrow E[g(X)] \geq \underbrace{g[E(X)]}_{\text{ged}} + c[E(X) - E(X)] = g[E(X)]$$

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

Proposition  $Y \geq 0$  &  $E(Y) = 0 \Rightarrow P(Y=0) = 1$  (ie  $Y \stackrel{a.s.}{=} 0$ )

The proof uses Boole & Markov.

rvec  
mean  $\underline{\mu} = E(\underline{X}) = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_k) \end{pmatrix}$ ;  $\Sigma = \text{Var}(X) = \left\{ \text{cov}(X_i, X_j) \right\}_{i,j=1}^k$

Lemma (i)  $E(A \tilde{X} + \tilde{b}) = A E(\tilde{X}) + \tilde{b}$

(ii)  $\text{Var}(A \tilde{X} + \tilde{b}) = A \Sigma A'$

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

Consequence  $\underline{c}' \Sigma \underline{c} \geq 0, \forall \underline{c}$

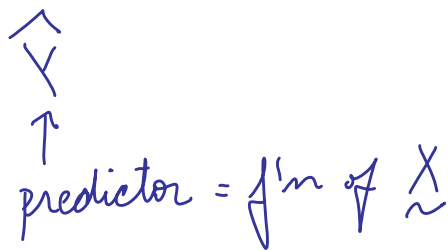
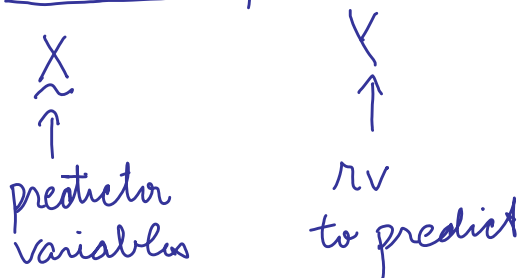
Assume  $\underline{c}' \Sigma \underline{c} > 0, \forall \underline{c} \neq \underline{0}$

$\Sigma$  is positive definite ( $\Sigma > 0$ )

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

triangular diagonal (elements)

Linear prediction



Assume 0 means & only consider linear predictors

$\hat{Y} = \underline{a}' \tilde{X}$

Set  $\underset{\sim}{\Sigma}_{XX} = E(\underset{\sim}{X} \underset{\sim}{X}') (= \text{Var}(\underset{\sim}{X}) \text{ if } 0 \text{ means})$

$$\underset{\sim}{\Sigma}_{YX} = E(Y \underset{\sim}{X})$$

product moment matrix

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

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

$$= E(Y^2) + \underset{\sim}{a}' \underset{\sim}{\Sigma}_{XX} \underset{\sim}{a} - 2 \underset{\sim}{a}' \underset{\sim}{\Sigma}_{YX}$$

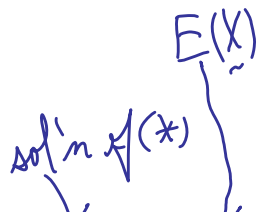
$$\frac{\partial \text{MSE}(\underset{\sim}{a})}{\partial \underset{\sim}{a}'} = \underset{\sim}{0}'$$

yields the equation

$$\underset{\sim}{\Sigma}_{XX} \underset{\sim}{a} = \underset{\sim}{\Sigma}_{YX} \quad (*)$$

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

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



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

### Lebesgue Monotone Convergence Theorem (MCT)

$$0 \leq X_n \uparrow X \Rightarrow E(X_n) \rightarrow E(X)$$

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

### Lebesgue Dominated Convergence Theorem (DCT)

$$X_n \rightarrow X + |X_n| \leq W, \forall n \quad \text{with } E(W) < \infty$$

then  $E(X_n) \rightarrow E(X)$ .

Remark In the DCT we may in fact show  $E(|X_n - X|) \rightarrow 0$ , which is called convergence in  $L_1$ . A notation for this is  $X_n \xrightarrow{L_1} X$

Try these 2 -  $X_n \xrightarrow{ms} X \Rightarrow X_n \xrightarrow{L_1} X \Rightarrow E(X_n) \rightarrow E(X)$   
- derive Chebyshev from Markov