Contents

Conditional Probability

Marginalisation

$P(E \cap F_i) = P(F_i \mid E)P(E)$ so,

$P(E \cap F_1) + P(E \cap F_2) + \ldots P(E \cap F_n)$

$= P(F_1 \mid E)P(E) + P(F_2 \mid E)P(E) + \ldots P(F_n \mid E)P(E)$

$= (P(F_1 \mid E) + P(F_2 \mid E) + \ldots P(F_n \mid E))P(E)$

$= P(E)$

since $P(F_1 \mid E) + P(F_2 \mid E) + \ldots P(F_n \mid E) = P(S \mid E) = 1$

Random Variables

Marginalisation

$P(X = x \text{ and } Y = y_i) = P(Y = y_i \mid X = x)P(X = x)$ so,

$\sum_{i=1}^{m} P(X = x \text{ and } Y = y_i) = \sum_{i=1}^{m} P(Y = y_i \mid X = x)P(X = x)$

$= P(X = x) \sum_{i=1}^{m} P(Y = y_i \mid X = x)$

$= P(X = x)$

since $\sum_{i=1}^{m} P(Y = y_i \mid X = x) = 1$

Expected Value

Linearity

Random variable X takes values $x_1, x_2, \ldots, x_n$ so,

$E[aX + b] = \sum_{i=1}^{n}(ax_i + b)P(X = x_i)$

$= \sum_{i=1}^{n} ax_i P(X = x_i) + \sum_{i=1}^{n} bP(X = x_i)$

$= a \sum_{i=1}^{n} x_i P(X = x_i) + b \sum_{i=1}^{n} P(X = x_i)$

$= aE[X] + b$

Two Random Variables

$E[aX + bY] = \sum_x \sum_y (ax + by)P(X = x \cap Y = y)$

$= a \sum_x \sum_y xP(X = x \cap Y = y) + b \sum_x \sum_y yP(X = x \cap Y = y)$

$= a \sum_x xP(X = x) + b \sum_y yP(Y = y)$

$= aE[X] + bE[Y]$

since $\sum_y P(X = x \cap Y = y) = P(X)$

Independent Random Variables

$E[XY] = \sum_x \sum_y xyP(X = x \text{ and } Y = y)$

$= \sum_x \sum_y xyP(X = x)P(Y = y)$

$= \sum_x xP(X = x) \sum_y yP(Y = y)$

$= E[X]E[Y]$

Variance

$Var(X) = \sum_{i=1}^{n}(x_i - \mu)^2 P(x_i)$

$= \sum_{i=1}^{n}(x_i^2 - 2\mu x_i + \mu^2) P(x_i)$

$= \sum_{i=1}^{n} x_i^2 P(x_i) - 2\sum_{i=1}^{n} x_i P(x_i)\mu + \mu^2 \sum_{i=1}^{n} P(x_i)$

$= E[X^2] - 2\mu^2 + \mu^2$

$= E[X^2] - (E[X])^2$

Non Linearity

$Var(aX + b) = E[(aX + b)^2] - E[aX + b]^2$

$= E[a^2X^2 + 2abX + b^2] - (aE[X] + b)^2$

$= a^2E[X^2] + 2abE[X] + b^2 - a^2E[X]^2 - 2abE[X] - b^2$

$= a^2E[X^2] - a^2E[X]^2$

$= a^2(E[X^2] - E[X]^2)$

$= a^2Var(X)$

Indepedent Random Variables

$Var(X + Y) = E[(X + Y)^2] - E[X + Y]^2$

$= E[X^2 + 2XY + Y^2] - (E[X] + E[Y])^2$

$= E[X^2] + 2E[XY] + E[Y^2] - E[X]^2 - 2E[X]E[Y] - E[Y]^2$

$= E[X^2] - E[X]^2 + E[Y^2] - E[Y]^2 + 2E[XY] - 2E[X]E[Y]$

$= E[X^2] - E[X]^2 + E[Y^2] - E[Y]^2$

$= Var(X) + Var(Y)$

Covariance

$Cov(X, Y) = E[(X - \mu_x)(Y - \mu_y)]$

$Cov(X, Y) = E[XY - X\mu_y - Y\mu_x + \mu_x\mu_y]$

$= E[XY] - E[X]\mu_y - E[Y]\mu_x + \mu_x\mu_y$

$= E[XY] - \mu_x\mu_y - \mu_x\mu_y + \mu_x\mu_y$

$= E[XY] - \mu_x\mu_y$

$= E[XY] - E[X]E[Y]$

Inequalities

Markov

Let indicator $I_a(X) = 1$ if $X \geq a$ and $I_a(X) = 0$. Then $aI_a(X) \leq X$, i.e. $I_a(X) \leq \frac{X}{a}$

$E(I_a(X)) \leq E(\frac{X}{a}) = \frac{E(X)}{a}$

$E(I_a(X)) = P(X \geq a) \leq \frac{E(X)}{a}$

Chebyshev

Since $(X - \mu)^2$ is a non-negative random variable we can apply Markov's inequality with $a = k^2$ to get

$P((X - \mu)^2 \geq k^2) \leq \frac{E((X-\mu)^2)}{k^2} = \frac{\sigma^2}{k^2}$

Note that $(X - \mu)^2 \geq k^2 \Leftrightarrow \mid X - \mu \mid \geq k$, so

$P(\mid X - \mu \mid \geq k) \leq \frac{\sigma^2}{k^2}$

Chernoff

$P(X \geq a) = P(e^{tX} \geq e^{ta})$ for $t > 0$

By Markov's inequality:

$P(X \geq a) = P(e^{tX} \geq e^{ta}) \leq \frac{E(e^{tX})}{e^{ta}} = e^{-ta}E(e^{tX})$

This holds for all $t > 0$, so might as well choose the one that minimises it.

Weak Law of Large Numbers

$E(\bar{X}) = E(\frac{1}{N}\sum_{k=1}^{N} X_k) = \frac{1}{N}\sum_{k=1}^{N} E(X_k) = \mu$

$var(\bar{X}) = var(\frac{1}{N}\sum_{k=1}^{N} X_k) = \frac{N\sigma^2}{N^2} = \frac{\sigma^2}{N}$