Expectation of Jointly Distributed Random Variables

Prop. 1 Let X and Y be jointly distributed.

(1) In the discrete case,

$$\boxed{E[g(X,Y)] = \sum_{\substack{x \\ \in \\ R(X)}} \sum_{\substack{y \\ \in \\ R(Y)}} g(x,y)\,p(x,y)}$$

(2) In the continuous case,

$$\boxed{E[g(X,Y)] = \iint_{\mathbb{R}^2} g(x,y)\,f(x,y)\,dA}$$

for any function $g: \mathbb{R}^2 \to \mathbb{R}$.

pf: (1) For the discrete case, we use Prop. 4.1,

$$E[g(X)] = \sum_{x \in R(X)} g(x)\,p(x)$$

Letting $Z = g(X, Y)$,

$$E[Z] \overset{\text{Prop. 4.1}}{=} \sum_{z \in R(g(X,Y))} \overset{g(x,y)}{\overset{''}{z}}\, p_Z(\overset{''}{z})$$

$$= \sum_{z \in R(g(X,Y))} \underset{}{\overset{=g(x,y)}{z}}\, \overbrace{P(Z=z)}^{= P(\overbrace{Z = g(X,Y) = z}^{\text{i.e. } (X,Y) \in g^{-1}(z)})}$$

$$= \sum_{\substack{(x,y) \in g^{-1}(z) \\ \cap R(X)\times R(Y)}} P(X=x, Y=$$

$$= \sum_{z \in R(g(X,Y))} g(x,y) \cdot \sum_{\substack{(x,y) \in g^{-1}(z) \\ \cap (R(X) \times R(Y))}} P(X=x, Y=y)$$

$$= \sum_{x \in R(X)} \sum_{y \in R(Y)} g(x,y)\, p(x,y)$$

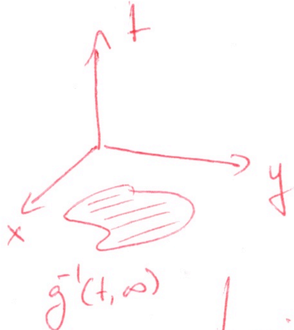
(2) For the continuous case, we use Exercises 5.2 – 5.3 (which complete Prop. 2.1):

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$$

Apply this to $z = g(X,Y) : S^2 \to \mathbb{R}$,

$$E[g(X,Y)] = \int_0^\infty P(g(X,Y)) > t)\, dt$$

$$- \int_0^\infty P(g(X,Y) < -t)\, dt$$



horizontal slicing

$$= \int_0^\infty \left(\iint_{g^{-1}(t,\infty)} f(x,y)\, dx\, dy \right) dt$$

$$- \int_0^\infty \left(\iint_{g^{-1}(-\infty,-t)} f(x,y)\, dx\, dy \right) dt$$

$$= \int_0^\infty \int_0^\infty \int_0^{g(x,y)} f(x,y)\, dt\, dx\, dy$$

$$- \int_{-\infty}^0 \int_{-\infty}^0 \int_{-g(x,y)}^0 f(x,y)\, dt\, dx\, dy$$

$$= \iint_{\mathbb{R}^2} g(x,y) f(x,y)\, dx\, dy$$

Ex. 1 $\quad E[X+Y] = \iint_{\mathbb{R}^2} (x+y)\, f(x,y)\, dA$

$$= \iint_{\mathbb{R}^2} x\, f(x,y)\, dA + \iint_{\mathbb{R}^2} y\, f(x,y)\, dA$$

$$= \int_{\mathbb{R}} x\, f_X(x)\, dx + \int_{\mathbb{R}} y\, f_Y(y)\, dy$$

(since $\int_{\mathbb{R}} f(x,y)\, dy = f_X(x)$

& $\int_{\mathbb{R}} f(x,y)\, dx = f_Y(y)$)

$$= E[X] + E[Y]$$

Ex 2

$$X \ge Y \iff X - Y \ge 0$$

$$\Longrightarrow E[X-Y] = E[X] - E[Y] \ge 0$$

$$\Longrightarrow E[X] \ge E[Y]$$

since $Z \ge 0 \Rightarrow E[Z] = \int_0^\infty \underbrace{z f(z)}_{\ge 0}\, dz \ge 0.$

ex. Sample mean of independent random variables $X_1, \dots, X_n$:

$$\boxed{\bar{X} \stackrel{\text{def}}{=} \sum_{i=1}^{n} \frac{X_i}{n}} \quad \text{(sample mean)}$$

$$\Longrightarrow \boxed{E[\bar{X}] = \frac{1}{n} \sum_{i=1}^{n} E[X_i]}$$

If all $E[X_i] = \mu$, then $E[\bar{X}] = \mu$ too!

ex. Let $X_i : \{\Sigma, F\} \longrightarrow \{0,1\}$ be a Bernoulli random variable, & let

$$\boxed{X = \sum_{i=1}^{n} X_i}$$

be the resulting binomial random variable.

Then,

$$\boxed{E[X] = \sum_{i=1}^{n} E[X_i] = np}$$

since $E[X_i] = 0 \cdot (1-p) + 1 \cdot p = p$.

Prop. 1 (Boole's Inequality)

Let $A_1, \ldots, A_n \in \mathcal{A}(S)$ be events, and let X_i be the indicator functions

$$\boxed{X_i = \begin{cases} 1, & \text{if } A_i \text{ occurs} \\ 0, & \text{else} \end{cases}}$$

i.e. $X_i(s) = 1$ if $s \in A_i$ & $X_i(s) = 0$ if $s \notin A_i$.

Define

$$\boxed{X \stackrel{def}{=} \sum_{i=1}^{n} X_i}$$ (counts # of events A_i which occur)

and observe that if

$$\boxed{Y \stackrel{def}{=} \begin{cases} 1, & \text{if } X \geq 1 \\ 0, & \text{else} \end{cases}}$$

then $Y \leq X \Rightarrow$

$$E[Y] \leq E[X] = \sum_{i=1}^{n} E[X_i]$$

i.e.

$$\boxed{P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i)}$$

Prop. 2 (Inclusion-Exclusion via Random Variables' Expectations)

Let $A_1, \dots, A_n \in \mathcal{A}(S)$ be events. Then

$$\boxed{P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \cdots + (-1)^{n+1} P(A_1 \cap \cdots \cap A_n)}$$

pf: Let X_i be the indicator function of A_i,

$$X_i(s) = \begin{cases} 1, & \text{if } s \in A_i \\ 0, & \text{else} \end{cases}$$

and consider the random variable

$$\Big(1 - \underbrace{\prod_{i=1}^n (1 - X_i)}_{\text{FOIL this out}}\Big)(s) = \begin{cases} 1, & \text{if } s \in \bigcup_{i=1}^n A_i \\ 0, & \text{else} \end{cases}$$

Then,

$$P\left(\bigcup_{i=1}^n A_i\right) = E[1 - \prod_{i=1}^n (1 - X_i)] = E\Big[\sum_{i=1}^n X_i - \sum_{i<j} X_i X_j + \cdots + (-1)^{n+1} X_1 \cdots X_n\Big]$$

$$= \sum_{i=1}^{n} E[X_i] - \sum_{i<j} E[X_i X_j]$$

$$+ \cdots + (-1) E[X_1 \cdots X_n]$$

$$= \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j)$$

$$+ \cdots + (-1) P(A_1 \cap \cdots \cap A_n)$$

QED

Moments of Random Variables

<u>Def. 1</u> For any $n \in \mathbb{N}$, the <u>nth moment</u> of a random variable X is defined to be

$$\boxed{E[X^n]} = \sum_{x \in R(X)} x^n p(x) \quad (\underline{\text{discrete}})$$

OR

$$\int_{\mathbb{R}} x^n f(x)\, dx \quad (\underline{\text{continuous}})$$

The <u>first moment</u> of X is its <u>expected value</u> $E[X]$, or <u>mean</u> μ.

<u>Def. 2</u> The <u>nth central moment</u> of X is the nth moment of $(X-\mu) = (X - E[X])$, the 2nd being

$$\underbrace{E[(X-\mu)^2]}_{=Var(X)} = \underbrace{E[X^2]}_{\text{2nd moment}} - \underbrace{E[X]}_{\text{1st moment}}{}^2 :$$

$$\underbrace{\boxed{E[(X-\mu)^n]}}_{\text{nth central moment}} = \sum_{x \in R(X)} x^n p(x) \quad \text{OR} \quad \int_{\mathbb{R}} x^n f(x)\, dx$$

The 3rd central moment $E[(X-\mu)^3]$ is called skewness, and measures the lopsidedness or lack of symmetry about μ (left-skewed distributions, i.e. f with long left tail, will have negative skewness, those which are right-skewed will have positive skewness).

The 4th central moment $E[(X-\mu)^4]$ is called kurtosis, and measures the heaviness of the tail of the distribution as compared to the normal distribution of the same variance.

Ex. Moments of the Number of Events Occurring

Consider the second term $\sum_{i<j} X_i X_j$ in that previous sum. The $X_i X_j = \begin{cases} 1, & \text{if } A_i \cap A_j \text{ occurs} \\ 0, & \text{if not} \end{cases}$

so they count the number of pairwise intersections $A_i \cap A_j$ occurring.

On the other hand, $X = \sum_{i=1}^{n} X_i$ counts the total number of events occuring, so we can say, by analogy with counting

$$\boxed{\binom{X}{2} \equiv \frac{X(X-1)}{2} = \sum_{i<j} X_i X_j}$$

$$\Longrightarrow E\left[\frac{X(X-1)}{2}\right] \equiv E\left[\binom{X}{2}\right] = \sum_{i<j} E[X_i X_j]$$

$$\Longrightarrow E[X^2] - E[X] = 2\sum_{i<j} E[X_i X_j]$$

2nd moment

$$= 2\sum_{i<j} P(A_i \cap A_j)$$

1st moment

$$= \mu = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} P(A_i)$$

Therefore, since the 1st moment of X is its expected value/mean,

$$\boxed{\begin{aligned}\mu = E[X] &= \sum_{i=1}^{n} E[X_i] \\ &= \sum_{i=1}^{n} P(A_i)\end{aligned}}$$

its 2nd moment is given by

$$\boxed{\begin{aligned}E[X^2] &= E[X] + 2\sum_{i<j} E[X_i X_j] \\ &= \sum_{i=1}^{n} P(A_i) + \sum_{i<j} P(A_i \cap A_j)\end{aligned}}$$

It's 2nd central moment, or variance, therefore, is

$$\boxed{\begin{aligned}Var(X) &= E[(X-\mu)^2] \\ &= E[X^2] - E[X]^2 \\ &= \sum_{i=1}^{n} P(A_i) + \sum_{i<j} P(A_i \cap A_j) \\ &\quad - \left(\sum_{i=1}^{n} P(A_i)\right)^2\end{aligned}}$$

Since by the same argument we arrive at

$$\binom{X}{k} \overset{\text{def}}{=} \frac{X(X-1)\cdots(X-k+1)}{k!}$$

$$= \sum_{i_1<\cdots<i_k} X_{i_1}X_{i_2}\cdots X_{i_k}$$

we conclude that

$$E\left[\binom{X}{k}\right] = \sum_{i_1<\cdots<i_k} E[X_{i_1}\cdots X_{i_k}]$$

$$= \sum_{i_1<\cdots<i_k} P(A_{i_1}\cap\cdots\cap A_{i_k})$$

For example, $X(X-1)(X-2) = (X^2-X)(X-2)$

$$= X^3 - 3X^2 + 2X$$

$\Longrightarrow$

$$E[X^3-3X^2+2X] = 3! \sum_{i_1<i_2<i_3} P(A_{i_1}\cap A_{i_2}\cap A_{i_3})$$

$\Longrightarrow$

$$E[X^3] = 3E[X^2] - 2E[X] + 6\sum_{i_1<i_2<i_3} P(A_{i_1}\cap A_{i_2}\cap A_{i_3})$$

$$= \sum_{i=1}^{n} P(A_i) + 3\sum_{i<j} P(A_i\cap A_j) + 6\sum_{i<j<k} P(A_i\cap A_j\cap A_k)$$

etc.