3.4 Indicator method (source formulas)

$X = \sum_{i=1}^{n} 1_{A_i}$, $X^2 = \sum_{i=1}^{n} 1_{A_i} + 2\sum_{i<j} 1_{A_i A_j}$

(1) $E(X) = \sum_{i=1}^{n} P(A_i)$, $E(X^2) = \sum_{i=1}^{n} P(A_i) + 2\sum_{i<j} P(A_i A_j)$

$Var(X) = E(X^2) - (E(X))^2$

(2) $Var(X) = \sum_{i=1}^{n} Var(1_{A_i}) + 2\sum_{i<j} Cov(1_{A_i}, 1_{A_j})$, where

$Var(1_{A_i}) = P(A_i) - P(A_i)^2$

$Cov(1_{A_i}, 1_{A_j}) = E(1_{A_i} 1_{A_j}) - E(1_{A_i})E(1_{A_j}) = P(A_i A_j) - P(A_i)P(A_j)$

This week: 3.7 Conditional expectation

3.8 Sums of r.v.

3.5 Examples

3.7 Conditional expectations

Def. Conditional pmf of Y given $X = x$, is the pmf

$f(y|x) = f_{Y|X}(y|x) = P(Y=y \mid X=x) = \dfrac{P(X=x, Y=y)}{P(X=x)}$

$= \dfrac{f(x,y)}{f_X(x)}$ if $f_X(x) = P(X=x) > 0$, where

$f(x,y)$ is joint pmf of (X,Y)

Some formulas:

a) $f(y|x) = \dfrac{f(x,y)}{f_X(x)}$

b) $f(x,y) = f(y|x)\, f_X(x)$ (multiplication rule)

c) $f_Y(y) = \sum_x f(y|x)\, f_X(x)$ (total probability law)

d) $f(x|y) = \dfrac{f(y|x)\, f_X(x)}{f_Y(y)}$ (Bayes formula)

<u>Def</u>. Conditional expectation of Y given $X=x$ is the expectation

$$E(Y|X=x) = \sum_y y\, P(Y=y|X=x) = \sum_y y\, f(y|x).$$

<u>Def</u>. For an event A we define

$$E(Y|A) = \sum_y y\, P(Y=y|A)$$

<u>Def</u>. $E(Y|X)$ is discrete r.v. defined as a function of X.

$E(Y|X) = h(X)$, where $h(x) = E(Y|X=x)$, $-\infty < x < \infty$.

<u>Properties of $E(Y|X)$</u>

1. <u>$E(Y)$ can be computed using conditioning</u>:

(1) $$E(Y) = E[E(Y|X)] = \sum_x \underbrace{E(Y|X=x)}_{h(x)} P(X=x)$$

Also, if $\Omega = \bigcup_i A_i$, and A_i are disjoint, then

$$E(Y) = \sum_i E(Y|A_i)\, P(A_i).$$

Note (1) is equivalent to

(1') $E(Y) = E[h(X)]$, where $h(x) = E(Y|X=x)$.

Proof of $\boxed{(1)\quad E(Y) = E[E(Y|X)] = \sum_x E(Y|X=x)\, P(X=x)}$

Indeed, $E(Y) = \sum_y y\, P(Y=y) = \sum_y y \sum_x P(Y=y|X=x)\, P(X=x)$

$= \sum_x \left(\sum_y y\, P(Y=y|X=x)\right) P(X=x) = \sum_x E(Y|X=x)\, P(X=x).$

Application of $\boxed{E(Y) = \sum_x E(Y|X=x)\, P(X=x)} = E[E(Y|X)]$

Ex 1. Population is divided into r groups whose proportions are $p_1, \dots, p_r$ ($p_1 + \dots + p_r = 1$). Average weight of ith group member is w_i.

What is expected weight of a randomly chosen member of the population?

Answer. Let W be weight of a member

X be the group of a member

We know $P(X=i) = p_i$, $i = 1, \dots, r$

$E(W|X=i) = w_i$, $i = 1, \dots, r$.

$$E(W) = \sum_{i=1}^{r} E(W|X=i)\, P(X=i) = \sum_{i=1}^{r} w_i p_i$$

Application of (1'). $E(Y) = E[h(X)]$, where $h(x) = E(Y|X=x)$.

Application of (1'). $E(Y) = E[h(X)]$, where $h(x)$ $E(Y|X=x)$.

Ex2. Number N of fish J catches is Poisson(λ). Let Y be the number of H's in N tosses of a coin with $P(H)) = p$.

(i) Find $E(Y|N=n)$, $E(Y|N)$, $E(Y)$.

Answer. We know: a) Given $N=n$, Y is binomial(n,p):

$$P(Y=k|N=n) = \binom{n}{k} p^k q^{n-k}, \quad k=0,1,\dots,n.$$

(In such a case, we say Y is binomial(N,p))

b) N is Poisson(λ)

Thus $E(Y|N=n) = np$, $E(Y|N) = pN$

$E(Y) = E(pN) = pE(N) = p\lambda$.

(ii) Find $E(Y^2|N)$.

Answer. $E(Y^2/N=n) = npq + (np)^2 = pqn + p^2n^2$

$E(Y^2/N) = pqN + p^2N^2$

Note $E(Y^2) = pq\,E(N) + p^2 E(N^2)$.

Remark 1. In Ex2, Y can be described as

a) Y is binomial(N,p): given $N=n$, Y is binomial(n,p)

b) Alternatively to a) Y can be modeled as

(3) $\boxed{Y = \sum_{i=1}^{N} X_i}$, where X_i are independent Bernoulli(p), independent of N.

According to (3), given $N=n$, $Y=\sum_{i=1}^{n} X_i$ is binomial (n,p).

Prediction property of $E(Y|X)$

2. $h(X) := E(Y|X)$ is the unique solution to the following minimization problem.

$$\boxed{\text{Find a function of } X \text{ (denoted } \ell(X)) \text{ so that } E\left[(Y-\ell(X))^2\right] \le E\left[(Y-g(X))^2\right] \text{ for any other } g(X)}$$

We use $g(X)$ to predict (estimate) Y;

$E\left[(Y-g(X))^2\right]$ is the mean square error of our estimate (prediction)

Answer. $\ell(X) = E(Y|X)$ is the best mean square estimate (prediction) of Y based on X.

Comment. In Ex 2., we found $E(Y|N) = pN$. Assume $p=\frac{1}{2}$, $N=10$. The best mean square estimate prediction of Y is $\frac{1}{2}\cdot 10 = 5$.

Why $\ell(X) = E(Y|X)$ is the best?