Probability 2 - Notes 3

The conditional distribution of a random variable X given an event B.

Let X be a random variable defined on the sample space S and B be an event in S. Denote $P(X=x|B) \equiv \frac{P(X=x \text{ and } B)}{P(B)}$ by $f_{X|B}(x)$. This is a probability mass function. We can therefore find the expectation of X conditional on B. $E[X|B] = \sum_x x f_{X|B}(x)$.

Example We toss a coin twice. Let X count the number of heads, so $X \sim Binomial\left(2, \frac{1}{2}\right)$, and let B_1 be the event that the first outcome is a head and B_2 be the event that the first outcome is a tail. Then $P(B_1) = P(\{HT, HH\}) = \frac{1}{2}$ and $P(B_2) = P(\{TH, TT\}) = \frac{1}{2}$.

Hence $P(X=0|B_1) = 0$, $P(X=1|B_1) = \frac{P(\{HT\}}{P(B_1)} = \frac{1}{2}$ and $P(X=2|B_1) = \frac{P(\{HH\}}{P(B_1)} = \frac{1}{2}$. Then $E[X|B_1] = \frac{3}{2}$.

Also $P(X=0|B_2) = \frac{P(\{TT\}}{P(B_2)} = \frac{1}{2}$, $P(X=1|B_2) = \frac{P(\{TH\}}{P(B_2)} = \frac{1}{2}$ and $P(X=2|B_2) = 0$. Therefore $E[X|B_2] = \frac{1}{2}$.

We can also obtain the conditional distribution of $X|B_1$ and $X|B_2$ by considering the implications of the experiment. If B_1 occurs then $X|B_1$ equals $1+Y$ where Y counts the number of heads in the second toss of the coin, so $Y \sim Bernoulli\left(\frac{1}{2}\right)$. If B_2 occurs then $X|B$ equals Y. Hence $E[X|B_1] = 1 + E[Y] = 1 + \frac{1}{2}$ and $E[X|B_2] = E[Y] = \frac{1}{2}$.

We will now look at a similar law to the law of total probability which is for expectations. This can be used to find the expected duration of the sequence of games (expected number of games played) for the gambler's ruin problem.

The law of total probability for expectations

From the law of total probability, if $B_1, ..., B_n$ partition S then for any possible value of x,

$$P(X=x) = \sum_{j=1}^{n} P(X=x|B_j)P(B_j) = \sum_{j=1}^{n} f_{X|B_j}(x)P(B_j).$$

Multiplying by x and summing we obtain the **Law of Total Probability for Expectations**

$$E[X] = \sum_{j=1}^{n} E[X|B_j]P(B_j)$$

Example. Consider the set-up for a geometric distribution. We have a sequence of independent trials of an experiment, with probability p of success at each trial. X counts the number of trials till the first success.

Let B_1 be the event that the first trial is a success and B_2 be the event that the first trial is a failure.

When B_1 occurs, X must equal 1. So $P(X = 1 \text{ and } B_1) = P(B_1)$ and $P(X = x \text{ and } B_1) = 0$ if $x > 1$. Hence the distribution of $X|B$ is concentrated at the single value 1 i.e. $X|B$ is identically equal to 1.

If B is the event that the first trial is a failure, then the number of trials until a success in the subsequent trials, Y, has the same distribution as X. We also have carried out the first trial. Hence $X|B$ is equal to $1+Y$ where Y has the same distribution as X.

Hence $E[X|B_1] = 1$ and $E[X|B_2] = 1+E[Y] = 1+E[X]$. Therefore

$$E[X] = E[X|B_1]P(B_1)+E[X|B_2]P(B_2) = p\times 1+q\times(1+E[X]).$$

Therefore $E[X] = \frac{1}{p}$.

The gambler's ruin problem, the expected duration of the game.

We use the same notation as before. The gambler plays a series of games starting with a stake of k units. He stops playing when he reaches either M or N units, where $M \leq k \leq N$. Let T_k be the random variable for the number of games played (the duration of the game). Then, if B_1 and B_2 are the events 'the gambler wins the first game' and 'the gambler loses the first game', the law of total probability for expectations is just

$$E[T_k] = E[T_k|B_1]P(B_1)+E[T_k|B_2]P(B_2)$$

If he wins the first game he has $k+1$ units so the distribution of T_k given B_1 has the same distribution as $1+T_{k+1}$ where T_{k+1} measures the duration of the game starting from $k+1$ units. Hence $E[T_k|B_1] = 1+E[T_{k+1}]$. Similarly $E[T_k|B_2] = 1+E[T_{k-1}]$. Let E_k denote $E[T_k]$. Then

$$E_k = p(1+E_{k+1})+q(1+E_{k-1})$$

and hence we obtain the difference equation

$$pE_{k+1}-E_k+qE_{k-1} = -1$$

Note that $E_M = E_N = 0$ since the gambling stops playing immediately. When $p \neq \frac{1}{2}$ a particular solution to this equation is $E_k = Ck$ where $C = \frac{1}{q-p}$. When $p = \frac{1}{2}$ a particular solution is $E_k = Ck^2$ where $C = -1$. Now, as for differential equations, the general solution to the particular difference equation is the particular solution just obtained plus the general solution to the general equation $pE_{k+1}-E_k+qE_{k-1} = 0$.

Case when $p \neq \frac{1}{2}$.

$$E_k = \frac{k}{q-p}+A+B\left(\frac{q}{p}\right)^k$$

Since $0 = E_M = \frac{M}{q-p} + A + B\left(\frac{q}{p}\right)^M$ and $0 = E_N = \frac{N}{q-p} + A + B\left(\frac{q}{p}\right)^N$, $B = \frac{(N-M)}{(q-p)\left(\left(\frac{q}{p}\right)^M - \left(\frac{q}{p}\right)^N\right)}$ and $A = -\frac{M}{(q-p)} - B\frac{\left(\frac{q}{p}\right)^M}{\left(\left(\frac{q}{p}\right)^M - \left(\frac{q}{p}\right)^N\right)}$. If we write E_k as $E_k(M,N)$ to explicitly include the boundaries we obtain

$$E_k(M,N) = \frac{(k-M)}{(q-p)} - \frac{(N-M)}{(q-p)} \frac{\left(\left(\frac{q}{p}\right)^k - \left(\frac{q}{p}\right)^M\right)}{\left(\left(\frac{q}{p}\right)^N - \left(\frac{q}{p}\right)^M\right)}$$

Case when $p = \frac{1}{2}$.

$$E_k = -k^2 + A + Bk$$

Since $0 = E_M = -M^2 + A + BM$ and $0 = E_N = -N^2 + A + BN$, $B = N + M$ and $A = -MN$. Hence writing E_k as $E_k(M,N)$ to explicitly include the boundaries

$$E_k(M,N) = (k-M)(N-k)$$

Conditional distribution of $X|Y$ where X and Y are random variables.

For any value y of Y for which $P(Y = y) > 0$ we can consider the conditional distribution of $X|Y = y$ and find the expectation and variance of X over this conditional distribution, $E[X|Y = y]$ and $Var(X|Y = y)$. Let $f_{X|Y}(x|y) = P(X = x|Y = y)$. Consider the function of Y which takes the value $E[X|Y = y]$ when $Y = y$. This is a random variable which we denote by $E[X|Y]$. Similarly we define $Var(X|Y)$ and $E[g(X)|Y]$ to be the functions of Y (so random variables) which take value $Var(X|Y = y)$ and $E[g(X)|Y]$ when $Y = y$.

Theorem. (i) $E[X] = E[E[X|Y]]$, (ii) $Var(X) = E[Var(X|Y)] + Var(E[X|Y])$ and (iii) $G_X(t) = E[E[t^X|N]]$.

Proof We show that $E[g(X)] = E[E[g(X)|Y]]$. Now

$$E[g(X)|Y = y] = \sum_{x=0}^{\infty} g(x) f_{X|Y}(x|y) = \sum_{x=0}^{\infty} g(x) \frac{P(X = x, Y = y)}{P(Y = y)}$$

$$\begin{aligned} E[E[g(X)|Y]] &= \textstyle\sum_{y=0}^{\infty} E[g(X)|Y = y] P(Y = y) \\ &= \textstyle\sum_{y=0}^{\infty} \sum_{x=0}^{\infty} g(x) \frac{P(X=x,Y=y)}{P(Y=y)} P(Y = y) \\ &= \textstyle\sum_{x=0}^{\infty} g(x) \sum_{y=0}^{\infty} P(X = x, Y = y) \\ &= \textstyle\sum_{x=0}^{\infty} g(x) P(X = x) = E[g(X)] \end{aligned}$$

(i) If we let $g(X) = X$ we immediately obtain $E[X] = E[E[X|Y]]$.

(ii) If we let $g(X) = X^2$ we obtain $E[X^2] = E[E[X^2|Y]]$.

Now $Var(X|Y)) = E[X^2|Y] - (E[X|Y])^2$ and hence

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

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

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

(iii) If we let $g(X) = t^X$ we obtain $G_X(t) = E[t^X] = E[E[t^X|N]]$.

Example Let $X \sim Binomial(n, p)$ and $Y \sim Binomial(m, p)$ where X and Y are independent. Then $R = X + Y \sim Binomial(n+m, p)$.

$$\begin{aligned} P(X = x|R = r) &= \frac{P(X=x,R=r)}{P(R=r)} = \frac{P(X=x,Y=r-x)}{P(R=r)} = \frac{P(X=x)P(Y=r-x)}{P(R=r)} \\ &= \frac{{}^nC_x p^x q^{n-x} \times {}^mC_{r-x} p^{r-x} q^{m-r+x}}{{}^{n+m}C_r p^r q^{n+m-r}} = \frac{{}^nC_x {}^mC_{r-x}}{{}^{n+m}C_r} \end{aligned}$$

Hence the conditional distribution of $X|R = r$ is hypergeometric. This provides the basis of the 2×2 contingency table test of equality of two binomial p parameters in statistics.

Example The number of spam messages Y in a day has Poisson distribution with parameter μ. Each spam message (independently) has probability p of not being detected by the spam filter. Let X be the number getting through the filter. Then $X|Y = y$ has Binomial distribution with parameters $n = y$ and p. Let $q = 1 - p$.

Hence $E[X|Y = y] = py$, $Var(X|Y = y) = pqy$ and $E[t^X|Y = y] = (pt+q)^y$ so that $E[X|Y] = pY$, $Var(X|Y) = pqY$ and $E[t^X|Y] = (pt+q)^Y$. Therefore:

$$E[X] = E[E[X|Y]] = E[pY] = pE[Y] = p\mu$$

$$Var(X) = E[Var(X|Y)] + Var(E[X|Y]) = E[pqY] + Var(pY) = p(1-p)\mu + p^2\mu = p\mu$$

$$G_X(t) = E[E[t^X|Y]] = E[(pt+q)^Y] = G_Y(pt+q) = e^{\mu((pt+q)-1)} = e^{p\mu(t-1)}$$

But this is the p.g.f. of a Poisson r.v. with parameter $\lambda = p\mu$. Hence by the uniqueness of the p.g.f., $X \sim Poisson(p\mu)$.

Random Sums.

Let $X_1, X_2, X_3, ...$ be a sequence of independent identically distributed random variables (i.i.d. random variables), each with the same distribution, each having common mean μ, variance σ^2 and p.g.f. $G_X(t)$. Consider the random sum $Y = \sum_{j=1}^{N} X_j$ where the number in the sum, N is also a random variable and is independent of the X_j. Then we can use our results for conditional expectations.

Since $E[Y|N=n] = E[\sum_{j=1}^{n} X_j] = \sum_{j=1}^{n} E[X_j] = n\mu$, we obtain the result that $E[Y] = E[E[Y|N]] = E[N\mu] = E[N]\mu$.

Similarly $Var(Y|N=n) = n\sigma^2$ so that

$$Var(Y) = E[Var(Y|N)] + Var(E[Y|N]) = E[N\sigma^2] + Var(N\mu) = \sigma^2 E[N] + \mu^2 Var(N)$$

Also we can obtain an expression for the p.g.f. of Y.

$$E[t^Y|N=n] = E\left[e^{\sum_{j=1}^{n} X_j}\right] = \prod_{j=1}^{n} G_{X_j}(t) = (G_X(t))^n$$

so that

$$G_Y(t) = E[E[t^Y|N]] = E\left[(G_X(t))^N\right] = G_N(G_X(t))$$

Example

Let X_j be the amount of money the j^{th} customer spends in a day in a shop. The $X's$ are i.i.d. random variables with mean 20 and variance 10. The number of customers per day N has Poisson distribution parameter 100. The total spend Y in the day is $Y = \sum_{j=1}^{N} X_j$. So $E[Y] = (20)(100) = 2000$ and $Var(Y) = (10)(100) + (20)^2(100) = 41000$.