Chapter 2. Discrete Distributions

Objectives

- Basic Concepts & Expectations
- Binomial, Poisson, Geometric, Negative Binomial, and Hypergeometric Distributions
- Introduction to the Maximum Likelihood Estimation
- Basic Bivariate Distributions; joint, marginal & conditional pdf

Discrete Distributions

The (theoretical) population mean, population variance, and the population sd (standard deviation) are:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i = \sum_x x f(x), \quad \sigma^2 = \sum_x (x-\mu)^2 f(x), \quad \text{and} \quad \sigma = \sqrt{\sigma^2}$$

The sample mean, sample variance, and the sample sd from a dataset:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}, \quad s = \sqrt{s^2}$$

or equivalently, from a frequency table:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} f_i x_i, \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{k} \left\{ f_i (x_i - \bar{x})^2 \right\}, \quad s = \sqrt{s^2}$$

Note also:

$$\begin{aligned}
\sigma^2 &= \sum_x (x-\mu)^2 f(x) \\
&= \sum_x \left(x^2 - 2\mu x + \mu^2\right) f(x) \\
&= \sum_x x^2 f(x) - 2\mu \sum_x x f(x) + \mu^2 \sum_x f(x) \\
&= \sum_x x^2 f(x) - \mu^2 \qquad \text{(Why?)}
\end{aligned}$$

$\sum_x x^2 f(x)$ is called the **2nd moment about the origin**.

Ex 1.

1. Find the mean and sd of the following observations: $x_1 = 3, x_2 = 1, x_3, = 2, x_4 = 6, x_5 = 3$.
2. Let $f(x) = x/6$, $x = 1, 2, 3$ be the pmf of X. Find the mean (μ) and the sd (σ) of X.

Ex 2. Let the random variable X = the number of rolls of a regular die until the "first" 5 or 6. The probability of rolling a 5 or 6 is 1/3, thus the pmf of X is written as:

$$f(x) = P(X = x) = \left(\frac{2}{3}\right)^{x-1}\left(\frac{1}{3}\right), \quad x = 1, 2, 3, , \ldots$$

Find (a) the mean (μ) and the sd (σ) of X.

Answer:

$$\mu = \sum_{x=1} xf(x) = 1\left(\frac{1}{3}\right) + 2\left(\frac{2}{3}\right)\left(\frac{1}{3}\right) + 3\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right) + \cdots$$

You will see quite a few of similar infinite sums of this kind. Here is how to find the answer easily. First, let's call $\frac{1}{3} = a$. Then,

$$\mu = 1a + 2\left(\frac{2}{3}\right)a + 3\left(\frac{2}{3}\right)^2 a + \cdots$$

Write one more line like

$$\frac{2}{3}\mu = \qquad \left(\frac{2}{3}\right)a + 2\left(\frac{2}{3}\right)^2 a + \cdots$$

Subtract, we get

$$\begin{aligned} \frac{1}{3}\mu &= a + \left(\frac{2}{3}\right)a + \left(\frac{2}{3}\right)^2 a + \cdots \\ &= \frac{a}{1 - 2/3} = 3a \\ \therefore \quad \mu &= 9a = 3 \end{aligned}$$

Remark Textbook introduces another way of handling this kind of infinite sum. First recall the Taylor series expansion:

$$f(x) = f(a) + \frac{f'(a)}{1!}x + \frac{f''(a)}{2!}x^2 + \cdots$$

Next, apply the Taylor series expansion to $f(x) = (1 - x)^{-2}$ around 0, we get

$$(1 - x)^{-2} = 1 + 2x + 3x^2 + \cdots$$

Now, notice that we have $3\mu = 1 + 2\left(\frac{2}{3}\right) + 3\left(\frac{2}{3}\right)^2 + \cdots$. The RHS is $(1 - x)^{-2}$, where $x = 2/3$. That is, $\mu = \frac{(1-2/3)^{-2}}{3} = 3$.

$$E\left(X^2\right) = \sum_{x=1} x^2 f(x) = 1^2\left(\frac{1}{3}\right) + 2^2\left(\frac{2}{3}\right)\left(\frac{1}{3}\right) + 3^2\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right) + \cdots$$

Rewrite as

$$3E\left(X^2\right) = 1^2 + 2^2\left(\frac{2}{3}\right) + 3^2\left(\frac{2}{3}\right)^2 + \cdots$$

Write one more line like before

$$3 \cdot \left(\frac{2}{3}\right) E\left(X^2\right) = \qquad 1^2\left(\frac{2}{3}\right) + 2^2\left(\frac{2}{3}\right)^2 + 3^2\left(\frac{2}{3}\right)^3 + \cdots$$

Subtract the last two equations, we get

$$E\left(X^2\right) = 1 + 3\left(\frac{2}{3}\right) + 5\left(\frac{2}{3}\right)^2 + 7\left(\frac{2}{3}\right)^3 + \cdots$$

Write one more line again like before

$$\left(\frac{2}{3}\right) E\left(X^2\right) = 1\left(\frac{2}{3}\right) + 3\left(\frac{2}{3}\right)^2 + 5\left(\frac{2}{3}\right)^3 + \cdots$$

Subtract the last equation from the one above, we get

$$\begin{aligned}\left(\frac{1}{3}\right) E\left(X^2\right) &= 1 + 2\left(\frac{2}{3}\right) + 2\left(\frac{2}{3}\right)^2 + 2\left(\frac{2}{3}\right)^3 + \cdots \\ &= 1 + 2\left(\frac{2/3}{1-2/3}\right)\end{aligned}$$

$$E\left(X^2\right) = 15$$

$$\therefore \quad \sigma^2 = E\left(X^2\right) - \mu^2 = 15 - 9 = 6$$

Basel Problem: $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$

We begin with the Taylor series expansion of $\sin(x)$

$$\sin(x) = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \frac{1}{7!}x^7 + \cdots$$

Divide both sides by x

$$\frac{\sin(x)}{x} = 1 - \frac{1}{3!}x^2 + \frac{1}{5!}x^4 - \frac{1}{7!}x^6 + \cdots$$

Notice that the roots of $\sin(x)$ are $\pm n\pi$ (i.e., $\sin(\pm n\pi) = 0$). The above expression must have factors like

$$\begin{aligned}
&\left(1-\frac{x}{\pi}\right)\left(1+\frac{x}{\pi}\right)\left(1-\frac{x}{2\pi}\right)\left(1+\frac{x}{2\pi}\right)\left(1-\frac{x}{3\pi}\right)\left(1+\frac{x}{3\pi}\right)\cdots \\
&= \left(1-\frac{x^2}{\pi^2}\right)\left(1-\frac{x^2}{4\pi^2}\right)\left(1-\frac{x^2}{9\pi^2}\right)\cdots \\
&= 1-\left(\frac{1}{\pi^2}+\frac{1}{4\pi^2}+\frac{1}{9\pi^2}+\cdots\right)x^2+\cdots \\
&= 1-\frac{1}{\pi^2}\sum_{n=1}^{\infty}\frac{1}{n^2}x^2+\cdots
\end{aligned}$$

Comparing the coefficients of x^2 term, we get

$$\frac{1}{3!} = \frac{1}{\pi^2}\sum_{n=1}^{\infty}\frac{1}{n^2}$$

$$\therefore \quad \sum_{n=1}^{\infty}\frac{1}{n^2} = \frac{\pi^2}{6}$$

Ex 3. Find a constant c so that $f(x) = \dfrac{c}{x^2}, \quad x = \pm 1, \pm 2, \ldots$ is a pmf.

Definition 1. *(Mathematical) Expectation*
Univariate case: $X \sim f(x)$ pdf for a continuous rv; pmf for a discrete rv.

$$E(X) = \begin{cases} \displaystyle\int_{-\infty}^{\infty} x f(x)\,dx \\ \displaystyle\sum_{x} x f(x) \end{cases}$$

Multivariate case: $X = (X_1, X_2, \ldots, X_n) \sim f(x_1, x_2, \ldots, x_n)$ pdf or pmf

$$E\left\{u\left(X_1, X_2, \ldots, X_n\right)\right\} = \begin{cases} \displaystyle\int \cdots \int u(x_1, \ldots, x_n) f(x_1, \ldots, x_n)\,dx_1 \ldots dx_n \\ \displaystyle\sum_{x} \cdots \sum_{x} u(x_1, \ldots, x_n) f(x_1, \ldots, x_n) \end{cases}$$

Properties 1. *If k is a constant $E(k) = k$.*
Properties 2. *If k is a constant and $v(\cdot)$ is a function, then $E\{kv(X)\} = kE\{v(X)\}$.*
This can be extended to $E\left\{\sum_{i=1}^{m} k_i v_i(X)\right\} = \sum_{i=1}^{m} k_i E\left\{v_i(X)\right\}$.

Proof.

$\square$

Ex 4. Let the random variable X have the pmf $f(x) = \left(\frac{1}{2}\right)^x$, $x = 1, 2, 3, \ldots$. Find μ and σ^2.
Answer: You can use the same methods shown before or use the Taylor series expansion of:

$$(1-x)^{-2} = 1 + 2x + 3x^2 + \cdots \quad \text{and} \quad (1-x)^{-3} = 1 + \frac{3\cdot 2}{2\cdot 1}x + \frac{4\cdot 3}{2\cdot 1}x^2 + \frac{5\cdot 4}{2\cdot 1}x^3 + \cdots$$

$$\begin{aligned}
\mu = E(X) = \sum_{x=1} x f(x) &= 1\left(\frac{1}{2}\right) + 2\left(\frac{1}{2}\right)^2 + 3\left(\frac{1}{2}\right)^3 + \cdots \\
&= \frac{1}{2}\left\{1 + 2\left(\frac{1}{2}\right) + 3\left(\frac{1}{2}\right)^2 + \cdots\right\} \\
&= \frac{1}{2}\left(1 - \frac{1}{2}\right)^{-2} = 2
\end{aligned}$$

$$\begin{aligned}
E\{X(X+1)\} = \sum_{x=1} x(x+1)f(x) &= 1\cdot 2\left(\frac{1}{2}\right) + 2\cdot 3\left(\frac{1}{2}\right)^2 + 3\cdot 4\left(\frac{1}{2}\right)^3 + \cdots \\
&= \left(1\cdot 2\cdot \frac{1}{2}\right)\left\{1 + \frac{3\cdot 2}{2\cdot 1}\left(\frac{1}{2}\right) + \frac{4\cdot 3}{2\cdot 1}\left(\frac{1}{2}\right)^2 + \cdots\right\} \\
&= \left(1 - \frac{1}{2}\right)^{-3} = 8
\end{aligned}$$

So, $E\left(X^2\right) = 6$ (Why?)

$$\therefore \quad \sigma^2 = E\left(X^2\right) - \{E(X)\}^2 = 6 - 4 = 2$$

Some special mathematical expectations

- Mean value of X:

$$E(X) = \mu = \begin{cases} \displaystyle\int x f(x)\,dx \\ \displaystyle\sum_x x f(x) \end{cases}$$

- Variance of X:

$$E(X-\mu)^2 = \sigma^2 = \begin{cases} \displaystyle\int (x-\mu)^2 f(x)\,dx \\ \displaystyle\sum_x (x-\mu)^2 f(x) \end{cases}$$

- Moment generating function (mgf) of X:

$$M(t) = \begin{cases} \displaystyle\int e^{tx} f(x)\,dx \\ \displaystyle\sum_x e^{tx} f(x) \end{cases}$$

Related facts

- $\sigma^2 = E\left(X^2\right) - \{E(X)\}^2$
- (sd) $\sigma = \sqrt{\sigma^2}$
- Not every distribution has an mgf. Suppose $M_X(t) = M_Y(t)$ for $|t| < h$, and some $h > 0$, then $P_X(\cdot) = P_Y(\cdot)$, i.e., $F_X(z) = F_Y(z), \quad \forall z$. This is called the "uniqueness of mgf", i.e., mgf uniquely determines the distribution.
- $M'(0) = E(X)$, $M''(0) = E\left(X^2\right)$, $\cdots$, $M^{(k)}(0) = E\left(X^k\right)$.

The last part is because

$$\frac{\partial}{\partial t} M(T) = \frac{\partial}{\partial t} E\left(e^{tX}\right) = E\left(Xe^{tX}\right)$$

Ex 5. {Cauchy distribution} X has pdf $f(x) = \dfrac{1}{\pi}\dfrac{1}{1+x^2}$, $-\infty < x < \infty$. Then, both $E(X)$ & $M_X(t)$ do NOT exist.
(Why?):

$$\lim_{\substack{a\to-\infty \\ b\to\ \infty}} \int_a^b x\,\frac{1}{\pi}\frac{1}{1+x^2}\,dx = ? \quad \text{(Does it exist?)}$$

$$\begin{aligned}
\lim_{b\to\infty}\int_0^b x\,\frac{1}{\pi}\frac{1}{1+x^2}\,dx &= \lim_{b\to\infty}\frac{1}{\pi}\frac{1}{2}\left\{\log\left(1+x^2\right)\right\}\Big|_0^b = +\infty \\
\lim_{a\to-\infty}\int_a^0 x\,\frac{1}{\pi}\frac{1}{1+x^2}\,dx &= \lim_{a\to-\infty}\frac{1}{\pi}\frac{1}{2}\left\{\log\left(1+x^2\right)\right\}\Big|_a^0 = +\infty
\end{aligned}$$

Binomial distribution

Definition 2. $X \sim \mathit{binomial}(n,p)$
X has a binomial distribution with parameters n & p $(n = 1, 2, \ldots, \quad 0 \le p \le 1)$
$\Leftrightarrow \quad f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \ldots, n$

$X \sim \mathit{binomial}(1,p)$ is particularly called the "Bernoulli" random variable, i.e., $P(X=1) = p = 1 - P(X=0)$

Properties 3. $X \sim \mathit{binomial}(n,p) \quad \Leftrightarrow \quad M_X(t) = \left(pe^t + q\right)^n, \quad q = 1-p.$

Proof.

$$(\Rightarrow) \quad M_X(t) = E\left(e^{tX}\right) = \sum_{x=0}^{n}\binom{n}{x}\left(pe^t\right)^x (1-p)^{n-x} = \left(pe^t + q\right)^n$$

$$(\Leftarrow) \quad \text{(Uniqueness of mgf)}$$

$\square$

Properties 4. *Representational definition of binomial* (n, p)

$$X \sim binomial(n,p) \quad \Leftrightarrow \quad X = \sum_{i=1}^{n} Z_i, \quad Z_1, \ldots, Z_n : iid \;\; binomial(1,p).$$

Proof. Begin with the mgf of $\sum Z_i$

$$\begin{aligned}
M_X(t) &= E\left(e^{t\sum Z_i}\right) \\
&= E\left(e^{tZ_1}e^{tZ_2}\cdots e^{tZ_n}\right) \\
&= \prod_1^n E\left(e^{tZ_i}\right) && \text{(Why?)} \\
&= \prod_1^n M_{Z_i}(t) && \text{(Why?)} \\
&= \prod_1^n \left(pe^t + q\right) && \text{(Why?)} \\
&= \left(pe^t + q\right)^n
\end{aligned}$$

$$\therefore \quad X \sim \text{binomial}\,(n,p) \quad \Leftrightarrow \quad X = \sum_{i=1}^{n} Z_i, \quad Z_1, \ldots, Z_n : \text{iid} \quad binomial(1,p)$$

The last part is by the uniqueness of mgf. □

Properties 5. *Mean & variance of binomial* (n, p)

$$X \sim binomial(n,p) \quad \Rightarrow \quad E(X) = np, \quad Var(X) = npq.$$

Proof.

- Easiest way: use the representational definition
- Proof by mgf: Try on your own using $M'(0) = E(X)$, $M''(0) = E\left(X^2\right)$
- Proof by pmf:

$$\begin{aligned}
E(X) &= \sum_{x=0}^{n} x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} p^x (1-p)^{n-x} && (\text{let } k = x-1) \\
&= np \sum_{k=0}^{n-1} \frac{(n-1)!}{k!(n-1-k)!} p^k (1-p)^{n-1-k} \\
&= np && \text{(Why?)}
\end{aligned}$$

$$
\begin{aligned}
E\{X(X-1)\} &= \sum_{x=0}^{n} x(x-1)\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x} \\
&= \sum_{x=2}^{n} \frac{n!}{(x-2)!(n-x)!}p^x(1-p)^{n-x} && (\text{let } k = x-2) \\
&= n(n-1)p^2 \sum_{k=0}^{n-2} \frac{(n-2)!}{k!(n-2-k)!}p^k(1-p)^{n-2-k} \\
&= n(n-1)p^2 && (\text{Why?})
\end{aligned}
$$

$$\therefore \quad \sigma^2 = E\{X(X-1)\} + E(X) - \{E(X)\}^2 = n(n-1)p^2 + np - (np)^2 = np(1-p)$$

□

Ex 6. {WLLN: Weak Law of Large Numbers} (Binomial case)

Chebyshev's Inequality: $P(|X-\mu| \geq k\sigma) \leq \dfrac{1}{k^2}$

Chebyshev's inequality came from the following:

$$\sigma^2 = E\left\{(X-\mu)^2\right\} = \sum_{x \in S}(x-\mu)^2 f(x) \ \geq \sum_{x \in A}(x-\mu)^2 f(x),$$

where $A = \{x : |x-\mu| \geq k\sigma\}$ for a positive constant k. This leads to

$$\sigma^2 \ \geq \sum_{x \in A}(x-\mu)^2 f(x) \ \geq \sum_{x \in A} k^2\sigma^2 f(x) = k^2\sigma^2 \sum_{x \in A} f(x) = k^2\sigma^2 P(X \in A)$$

□

Now, let $X_1, X_2, \ldots, X_n$ be an iid *binomial*$(1,p)$, $\hat{p} = \dfrac{1}{n}\displaystyle\sum_{1}^{n} X_i$ = sample success ratio. Consider

$$P\left(|\hat{p}-p| \geq \epsilon\right)$$

Note that $E\left(\hat{p}\right) = p$ and $\sigma\left(\hat{p}\right) = \sqrt{\frac{p(1-p)}{n}}$, so by plugging into the Chebyshev's inequality we get

$$P\left(|\hat{p}-p| \geq \epsilon\right) \leq \frac{p(1-p)}{\epsilon^2 n}$$

$$\therefore \quad \lim_{n\to\infty} P\left(|\hat{p}-p| \geq \epsilon\right) \leq \lim_{n\to\infty} \frac{p(1-p)}{\epsilon^2 n} = 0$$

This means that the probability that the sample success ratio $(\hat{p})$ is more than ϵ away from p goes to zero as n goes to ∞, and we say $(\hat{p})$ **converges in probability** to p.

□

Definition 3. *Cumulative distribution function (cdf) – univariate case*

$$F(x) = P\left(X \leq x\right)$$

Properties 6. *cdf* $F(x)$

1. *(monotonicity)* $a \leq b \qquad \rightarrow \qquad F(a) \leq F(b)$
2. $F(-\infty) = \lim_{x\to-\infty} F(x) = 0, \qquad F(+\infty) = \lim_{x\to\infty} F(x) = 1$
3. *(right continuity)* $\lim_{h\to 0^+} F(x+h) = F(x)$

A random variable X may not have a pdf or mgf, BUT it always has a cdf.

Definition 4. *Relationship with pdf (or pmf) when it exists*

$$F(x) = \begin{cases} \displaystyle\int_{-\infty}^{x} f(t)\,dt & \textit{(continuous case)} \\ \displaystyle\sum_{t\leq x} f(t) & \textit{(discrete case)} \end{cases}$$

$$f(x) = \begin{cases} \dfrac{\partial}{\partial x} F(x) \quad \textit{for } x \textit{ where } f \textit{ is continuous} & \textit{(continuous case)} \\ F(x) - F(x-) & \textit{(discrete case)} \end{cases}$$

Ex 7.

1. Find the cdf $F(x)$ of

$$f(x) = \begin{cases} \dfrac{x}{6} & \text{for } x = 1, 2, 3 \\ 0 & \text{otherwise} \end{cases}$$

2. Find the cdf $F(x)$ of

$$f(x) = \begin{cases} \dfrac{2}{x^3} & \text{for } x > 1 \\ 0 & \text{otherwise} \end{cases}$$

Answer:

$$F(x) = \begin{cases} 0 & x < 0 \\ 1/6 & 1 \leq x < 2 \\ 3/6 & 2 \leq x < 3 \\ 1 & x \geq 3 \end{cases}$$

$$F(x) = \begin{cases} \displaystyle\int_1^x \frac{2}{t^3}\,dt = 1 - \frac{1}{x^2}, & x > 1 \\ 0 & \text{otherwise} \end{cases}$$

Ex 8. Let $X \sim$ binomial (8, 0.65). Find (a) $P(X \le 5)$, (b) $P(X = 5)$, (c) $P(X \le 5) - P(X \le 4)$
Answer by R:

```
> pbinom(5,8,0.65)
[1] 0.5721863
> dbinom(5,8,0.65)
[1] 0.2785858
> pbinom(5,8,0.65)-pbinom(4,8,0.65)
[1] 0.2785858
```

Poisson distribution

Poisson approximation to the binomial distribution:

$$\binom{n}{x} p^x q^{n-x} \qquad \to \qquad \frac{e^{-\lambda}\lambda^x}{x!}, \qquad \text{as} \quad n \to \infty, \quad np = \lambda \text{ (fixed)}$$

There are different ways to show this. Here is an easy way from the textbook. We begin with

$$P(X = x) = \frac{n!}{x!(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

Next, let $n \to \infty$

$$\begin{aligned} P(X = x) &= \lim_{n\to\infty} \frac{n!}{x!(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\ &= \lim_{n\to\infty} \frac{n(n-1)\cdot(n-x+1)}{n^x} \frac{\lambda^x}{x!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x} \\ &= 1 \cdot \frac{\lambda^x}{x!} \cdot e^{-\lambda} \qquad \text{(Why?)} \\ &= \frac{e^{-\lambda}\lambda^x}{x!} \quad \sim \quad \text{Poisson with } \lambda \end{aligned}$$

Definition 5. $X \sim \textit{Poisson}\,(\lambda)$

X has a Poisson distribution with parameter $\lambda\,(>0)$ $\quad \Leftrightarrow \quad f(x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$

Properties 7. $X \sim \textit{Poisson}\,(\lambda) \quad \Leftrightarrow \quad M_X(t) = \exp\left\{\lambda\left(e^t - 1\right)\right\}$.

Proof.

$$M_X(t) = E\left(e^{tX}\right) = \sum_{x=0}^{n} e^{tx}\frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda}\sum_{x=0}^{n}\frac{\left(\lambda e^t\right)^x}{x!} = e^{-\lambda}e^{\left(\lambda e^t\right)} = e^{\left\{\lambda\left(e^t-1\right)\right\}}$$

□

Properties 8. *Mean & variance of Poisson* (λ)

$$X \sim Poisson\,(\lambda) \quad \Rightarrow \quad E(X) = \lambda, \quad Var(X) = \lambda.$$

Properties 9. *(Reproductive property)*

$X_1, \ldots, X_k$ *independent &* $X_i \sim Poisson\,(\lambda_i), \quad i = 1, 2\ldots, k,$
then $\sum X_i \sim Poisson\ (\sum \lambda_i)$

Proof. Begin with the mgf of $\sum Z_i$

$$\begin{aligned}
M_{\sum X_i}(t) &= E\left(e^{t\sum X_i}\right) \\
&= E\left(e^{tX_1}e^{tX_2}\cdots e^{tX_n}\right) \\
&= \prod_{1}^{n} E\left(e^{tX_i}\right) && \text{(Why?)} \\
&= \prod_{1}^{n} M_{X_i}(t) && \text{(Why?)} \\
&= \prod_{1}^{n} e^{\lambda_i\left(e^t-1\right)} \\
&= e^{\sum\lambda_i\left(e^t-1\right)} \\
&\sim \ \text{mgf of Poisson}\left(\sum \lambda_i\right)
\end{aligned}$$

$\therefore \sum X_i \sim$ Poisson $(\sum \lambda_i)$ by the uniqueness of mgf. □

Ex 9. Let $X \sim$ Poisson $(\lambda = 1)$. Find (a) $P\,(X \leq 2)$, (b) $P(X = 2)$, (c) $P\,(X \leq 2) - P\,(X \leq 1)$
Answer by R:

```
> ppois(2,1)
[1] 0.9196986
> dpois(2,1)
[1] 0.1839397
> ppois(2,1)-ppois(1,1)
[1] 0.1839397
```

Geometric distribution

Definition 6. $Y \sim Geometric\,(p) \quad \Leftrightarrow \quad f(y) = pq^y, \quad y = 0, 1, \ldots; q = 1-p$

$X_1, \ldots, X_n \sim$ iid binomial $(1, p)$, $Y = \#$ of "failures" before the first success.

Properties 10. $Y \sim Geometric\,(p) \quad \Leftrightarrow \quad M_Y(t) = p\left(1 - qe^t\right)^{-1}, \quad 1 - qe^t > 0.$
$\Leftrightarrow \quad E(Y) = \dfrac{q}{p}, \qquad Var(Y) = \dfrac{q}{p^2}.$

Proof.

$$M_Y(t) = E\left(e^{tY}\right) = \sum_{y=0}^{\infty} p\left(qe^t\right)^y = \frac{p}{(1 - qe^t)}, \quad 1 - qe^t > 0$$

□

One note: Textbook (page. 64) uses a slightly different definition. There, $X =$ the trial number on which the 1st "success" occurs and it's related by $Y = X - 1$, $(x = 1, 2, \ldots)$. According to this definition, we have

$$f(x) = pq^{x-1}\,(x = 1, 2, \ldots); E(X) = \frac{1}{p}, Var(X) = \frac{q}{p^2};\; M_X(t) = pe^t\left(1 - qe^t\right)^{-1}$$

Negative Binomial distribution

Definition 7. $Y \sim Negative\ Binomial\,(r, p)$

$$\Leftrightarrow \quad f(y) = \binom{y + r - 1}{r - 1} p^r q^y, \quad y = 0, 1, \ldots; q = 1 - p$$

$X_1, \ldots, X_n \sim$ iid binomial $(1, p)$, $Y = \#$ of "failures" before the rth success $(r \geq 1)$.

Properties 11. $Y \sim NB\,(r, p) \quad \Leftrightarrow \quad M_Y(t) = p^r\left(1 - qe^t\right)^{-r}, \quad 1 - qe^t > 0.$
$\Leftrightarrow \quad Y = \sum_1^r Z_i, \quad Z_1, \ldots, Z_r \sim iid\ geometric(p).$
$\Leftrightarrow \quad E(Y) = r\dfrac{q}{p}, \qquad Var(Y) = r\dfrac{q}{p^2}.$

Proof. Note first $\sum_{y=0}^{\infty} \binom{y + r - 1}{r - 1} q^y = p^{-r}$ (Why?). This is known as the "negative binomial expansion" and it's the Taylor series expansion of $f(q) = (1 - q)^{-r}$ as shown below

$$\begin{aligned} f(x) &= f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \cdots \\ (1 - q)^{-r} &= 1 \quad + \frac{r}{1!}q \quad + \frac{r(r + 1)}{2!}q^2 + \cdots \end{aligned}$$

Making use of this, we have

$$M_Y(t) = E\left(e^{tY}\right) = \sum_{y=0}^{\infty} \binom{y + r - 1}{r - 1} p^r \left(qe^t\right)^y = p^r\left(1 - qe^t\right)^{-r}, \quad 1 - qe^t > 0$$

Proof. ($mgf \quad \to \quad$ representational definition)

$$\begin{aligned}
M_{\sum Z_i}(t) &= E\left(e^{t\sum Z_i}\right), \quad \text{where} \sum_1^r Z_i \sim iid\ geometric\,(p) \\
&= \prod_1^r E\left(e^{tZ_i}\right) && \text{(Why?)} \\
&= \prod_1^r M_{Z_i}(t) && \text{(Why?)} \\
&= \prod_1^r \left\{p\left(1-qe^t\right)^{-1}\right\} \\
&= p^r\left(1-qe^t\right)^{-r} \\
&\sim \ \text{mgf of NB}\,(r,p)
\end{aligned}$$

$\therefore \sum Z_i \sim \text{NB}\,(r,p)$ by the uniqueness of mgf.

□

Another note: Textbook (page. 64) uses a slightly different definition. There, $X =$ the trial number on which the rth "success" occurs and it's related by $Y = X - r$, $(x = r, r+1, \ldots)$. Textbook calls Y has a "translated negative binomial" distribution. According to this definition, we have

$$f(x) = \binom{x-1}{r-1} p^r q^{x-r}\,(x = r, r+1, \ldots);\ \mu = \frac{r}{p}, \sigma^2 = \frac{rq}{p^2};\ M_X(t) = \left(pe^t\right)^r\left(1-qe^t\right)^{-r}$$

Hypergeometric distribution

Definition 8. $X \sim Hypergeometric\,(N_1, N_2, n)$

$$\Leftrightarrow \quad f(x) = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}}, \quad x \le n$$

In a box, there are N_1 "red" balls and N_2 "blue" balls, $X = \#$ of "red" balls.

Properties 12. $X \sim Hypergeom\,(N_1, N_2, n) \quad \Leftrightarrow \quad \mu = np$, $\sigma^2 = np(1-p)\left(\dfrac{N-n}{N-1}\right)$, $p = \dfrac{N_1}{N}$.

MLE: Maximum Likelihood Estimate

Definition 9. *Suppose $X_1, X_2, \ldots, X_n$ are random samples from the same underlying distribution (i.e., iid) with the pdf $f(x_i;\theta)$, then $\prod_{i=1}^n f(x_i;\theta)$ is called the **jpdf (joint pdf)** or the **likelihood***

function. *Furthermore, the value of the parameter $\hat{\theta}$ that maximizes the likelihood is called the mle (maximum likelihood estimator) of θ.*

Ex 10. $X_1, X_2, \ldots, X_n$: iid from binomial $(1, p)$. Find the mle of p.
Answer:

$$\begin{aligned} L(p) &= \prod_{i=1}^{n} f(x_i; p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}, \quad 0 < p < 1,\ x_i = 0 \text{ or } 1 \\ &= p^{\sum x_i}(1-p)^{n-\sum x_i} \\ \ln L(p) &= \sum x_i \ln p + \left(n - \sum x_i\right)\ln(1-p) \quad \text{(log-likelihood function)} \end{aligned}$$

Now, to find p that maximizes the log likelihood, differentiate this wrt p and set it equal to zero, we get

$$\begin{aligned} &\frac{\partial}{\partial p}\ln L(p) = \frac{\sum x_i}{p} - \frac{(n - \sum x_i)}{1-p} = 0 \\ &(1-p)\sum x_i - \left(n - \sum x_i\right)p = 0 \\ &\hat{p} = \frac{\sum x_i}{n} = \bar{x} \end{aligned}$$

To make sure that this indeed makes the log likelihood maximum, we can do the second derivative test. Here, we have

$$\frac{\partial^2}{\partial^2 p}\ln L(p) = -\frac{\sum x_i}{p^2} - \frac{(n - \sum x_i)}{(1-p)^2}$$

This is always < 0 regardless of $\hat{p}$, which means that our solution $\hat{p}$ is indeed the mle. **One note**: $\bar{X}$ is called the maximum likelihood **estimator**, and $\bar{x}$ is the maximum likelihood **estimate**.

Ex 11. $X_1, X_2, \ldots, X_n$: iid from Poisson (λ). Find the mle of λ.
Answer:

$$\begin{aligned} L(\lambda) &= \prod_{i=1}^{n} f(x_i; \lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{\lambda^{\sum x_i} e^{-n\lambda}}{x_1! x_2! \cdots x_n!} \\ \ln L(\lambda) &= -n\lambda + \sum x_i \ln \lambda - \ln(x_1! \cdots x_n!) \\ \frac{\partial}{\partial \lambda}\ln L(\lambda) &= -n + \frac{\sum x_i}{\lambda} = 0 \\ \hat{\lambda} &= \frac{\sum x_i}{n} = \bar{x} \end{aligned}$$

Now, the second derivative is $\frac{\partial^2}{\partial^2 \lambda}\ln L(\lambda) = -\frac{\sum x_i}{\lambda^2}$, and this is always < 0 regardless of $\hat{\lambda}$, which means $\hat{\lambda}$ is the mle.

Ex 12. $X_1, X_2, \ldots, X_n$: iid from discrete uniform for $x = 1, 2, ..., \theta$. Find the mle of θ.
Answer:

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} f(x_i;\theta) = \prod_{i=1}^{n} \frac{1}{\theta} = \left(\frac{1}{\theta}\right)^n, \qquad x_i = 1, \ldots, \theta \\ \ln L(\theta) &= -n \ln \theta \quad \Rightarrow \quad \frac{\partial}{\partial \theta} \ln L(\theta) = -\frac{n}{\theta} \\ \therefore\ \hat{\theta} &= \max_i (x_1, \ldots, x_n) \end{aligned}$$

This agrees with our intuition because in n observations of a discrete uniform random variable, the largest value should be taken as the upper bound.

Expected Values – Linear Functions of "Independent" Random variables

Ex 13. X_1, X_2: independent random samples from Poisson with $\lambda_1 = 2$, $\lambda_2 = 3$, respectively. Find (a) $P(X_1 = 3, X_2 = 4)$, (b) $P(X_1 + X_2 = 2)$.
Answer:

$$\begin{aligned} &P(X_1 = 3, X_2 = 4) = \left(\frac{2^3 e^{-2}}{3!}\right)\left(\frac{3^4 e^{-3}}{4!}\right) = \frac{9}{2} e^{-5} = 0.0303 \\ &P(X_1 + X_2 = 2) = P(0,2) + P(1,1) + P(2,0) \\ &= \left(\frac{2^0 e^{-2}}{0!}\right)\left(\frac{3^2 e^{-3}}{2!}\right) + \left(\frac{2^1 e^{-2}}{1!}\right)\left(\frac{3^1 e^{-3}}{1!}\right) + \left(\frac{2^2 e^{-2}}{2!}\right)\left(\frac{3^0 e^{-3}}{0!}\right) = 0.0842 \end{aligned}$$

Let $u(X_1, X_2)$ be a new random created as a function of two independent random variables X_1 and X_2, where X_1 and X_2 have pdf's $f_1(x_1), f_2(x_2)$, respectively. Then the expected value of $u(X_1, X_2)$ can be found by

$$E\{u(X_1, X_2)\} = \sum_{x_1}\sum_{x_2} u(x_1, x_2) f_{12}(x_1, x_2) = \sum_{x_1}\sum_{x_2} u(x_1, x_2) f_1(x_1) f_2(x_2)$$

The last part, where the joint pdf is written as a product of marginal pdf, is due to the independence of X_1 and X_2. Now, consider $Y = a_1X_1 + a_2X_2$, where a_1, a_2 are constants. We have

$$\begin{aligned} E(Y) = \mu_Y &= E(a_1X_1 + a_2X_2) = a_1E(X_1) + a_2E(X_2) = a_1\mu_1 + a_2\mu_2 \\ Var(Y) = \sigma_Y^2 &= E\left\{(a_1X_1 + a_2X_2 - a_1\mu_1 - a_2\mu_2)^2\right\} = E\left[\{a_1(X_1 - \mu_1) + a_2(X_2 - \mu_2)\}^2\right] \\ &= a_1^2 E\left\{(X_1 - \mu_1)^2\right\} + a_2^2 E\left\{(X_2 - \mu_2)^2\right\} + 2a_1a_2E\{(X_1 - \mu_1)(X_2 - \mu_2)\} \\ &= a_1^2 Var(X_1) + a_2^2 Var(X_2) + 2a_1a_2E(X_1 - \mu_1)E(X_2 - \mu_2) \\ &= a_1^2\sigma_1^2 + a_2^2\sigma_2^2 \end{aligned}$$

In general, let $Y = \sum_{i=1}^{n} a_iX_i$, where a_i's are constants and X_i's are **independent** random

samples with mean μ_i and variance σ_i^2, then

$$E(Y) = \mu_Y = E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E\left(X_i\right) = \sum_{i=1}^{n} a_i \mu_i$$
$$Var(Y) = \sigma_Y^2 = Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 Var\left(X_i\right) = \sum_{i=1}^{n} a_i^2 \sigma_i^2$$

One note: In case when $X_1, \ldots, X_n$ are not independent, we have

$$\sigma_Y^2 = \sum_{i=1}^{n} a_i^2 \sigma_i^2 + 2\sum\sum\nolimits_{i<j} a_i a_j \sigma_{ij}, \qquad \text{where } \sigma_{ij} = Cov(X_i, X_j)$$

Ex 14. Let $X_1 \sim binomial(n_1 = 100, p_1 = 1/2)$, $X_2 \sim binomial(n_2 = 48, p_2 = 1/4)$ and they are independent. Find the expected value and the variance of $Y = X_1 - X_2$.
Answer:

$$E(Y) = E\left(X_1 - X_2\right) = E\left(X_1\right) - E\left(X_2\right) = 50 - 12 = 38$$
$$Var(Y) = Var\left(X_1 - X_2\right) = Var\left(X_1\right) + Var\left(X_2\right) = 25 + 9 = 34$$

Definition 10. *$f\left(x_1, x_2\right)$ = joint pdf of X_1 and X_2.*

$$f_1\left(x_1\right) = \textit{marginal pdf of } X_1 = \sum_{x_2} f(x_1, x_2) \textit{ or } \int f(x_1, x_2) dx_2.$$

$$f\left(x_2 | x_1\right) = \frac{f\left(x_1, x_2\right)}{f_1\left(x_1\right)} = \textit{conditional pdf of } X_2 \textit{ given } X_1 = x_1.$$

Ex 15. Consider the following (discrete) joint pmf of X_1 and X_2.

	$X_1 = 1$	$X_1 = 2$	marginal
$X_2 = 1$	4/10	2/10	6/10
$X_2 = 2$	3/10	1/10	4/10
marginal	7/10	3/10	1

Find (a) $f(1, 2)$, (b) $f_1(1) \cdot f_2(2)$, (c) pmf of $Y = X_1 + X_2$, (d) $E(Y)$, and (e) $E\left(X_1 + X_2\right)$.

Answer:

$$f(1,2)=\frac{3}{10}; \qquad f_1(1)\cdot f_2(2)=\frac{7}{10}\cdot\frac{4}{10}=\frac{28}{100}; \qquad f(x_1,x_2)\neq f_1(x_1)f_2(x_2)$$

$$P(Y)=\begin{cases}4/10, & Y=2\\ 5/10, & Y=3\\ 1/10, & Y=4\end{cases}$$

$$E(Y)=2\left(\frac{4}{10}\right)+3\left(\frac{5}{10}\right)+4\left(\frac{1}{10}\right)=\frac{27}{10}$$

$$E\left(X_1+X_2\right)=E\left(X_1\right)+E\left(X_2\right)=1\left(\frac{7}{10}\right)+2\left(\frac{3}{10}\right)+1\left(\frac{6}{10}\right)+2\left(\frac{4}{10}\right)=\frac{27}{10}$$

Definition 11. $Cov\left(X_1,X_2\right)=\sigma_{12}=E\left(X_1-\mu_1\right)\cdot E\left(X_2-\mu_2\right)=E\left(X_1X_2\right)-\mu_1\mu_2$

$$\rho=Cor\left(X_1,X_2\right)=\frac{Cov(X_1,X_2)}{\sigma_1\sigma_2}=\frac{\sigma_{12}}{\sigma_1\sigma_2}.$$

Definition 12. *Conditional mean & conditional variance:*

$$E\left(X_2|x_1\right)=\mu_{X_2|x_1}=\begin{cases}\displaystyle\int_{-\infty}^{\infty}x_2f(x_2|x_1)dx_2\\ \displaystyle\sum_{x_2}x_2f\left(x_2|x_1\right)\end{cases}$$

$$Var\left(X_2|x_1\right)=\sigma^2_{X_2|x_1}=E\left[\left\{X_2-E\left(X_2|x_1\right)\right\}^2|x_1\right]=\begin{cases}\displaystyle\int_{-\infty}^{\infty}\left\{x_2-E\left(X_2|x_1\right)\right\}^2f(x_2|x_1)dx_2\\ \displaystyle\sum_{x_2}\left\{x_2-E\left(X_2|x_1\right)\right\}^2f\left(x_2|x_1\right)\end{cases}$$

The conditional variance can also be shown as

$$Var\left(X_2|x_1\right)=E\left(X_2^2|x_1\right)-\left\{E\left(X_2|x_1\right)\right\}^2$$

Ex 16. Let X_1 and X_2 have the joint pmf

$$f(x_1,x_2)=\frac{x_1+x_2}{21}, \quad x_1=1,2,3, \quad x_2=1,2$$

Here is the probability table for your information.

	$X_1=1$	$X_1=2$	$X_1=3$	marginal
$X_2=1$	2/21	3/21	4/21	9/21
$X_2=2$	3/21	4/21	5/21	12/21
marginal	5/21	7/21	9/21	1

Find (a) the marginal pmf's $f_1(x_1), f_2(x_2)$, (b) conditional pmf's $f(x_1|x_2), f(x_2|x_1)$, (c) conditional expectation $E\left(X_2|x_1\right)$, (d) conditional variance $Var\left(X_2|x_1\right)$, and (e) $P\left(X_1=2|X_2=2\right), E\left(X_2|x_1=3\right), Var\left(X_2|x_1=3\right)$.

Answer: Shown below are the "conditional" probabilities for your information. Check your calculation below with these values.

	$X_1=1$	$X_1=2$	$X_1=3$	conditional prob
$X_2=1$	2/9	3/9	4/9	$\Leftarrow$ $f(x_1\|x_2=1)$
$X_2=2$	3/12	4/12	5/12	$\Leftarrow$ $f(x_1\|x_2=2)$

	$X_1=1$	$X_1=2$	$X_1=3$
$X_2=1$	2/5	3/7	4/9
$X_2=2$	3/5	4/7	5/9
conditional prob	$f(x_2\|x_1=1)$	$f(x_2\|x_1=2)$	$f(x_2\|x_1=3)$

$$f_1(x_1)=\sum_{x_2}f(x_1,x_2)=\sum_{x_2=1}^{2}\frac{x_1+x_2}{21}=\frac{x_1+1}{21}+\frac{x_1+2}{21}=\frac{2x_1+3}{21},\quad x_1=1,2,3$$

$$f_2(x_2)=\sum_{x_1}f(x_1,x_2)=\sum_{x_1=1}^{3}\frac{x_1+x_2}{21}=\frac{1+x_2}{21}+\frac{2+x_2}{21}+\frac{3+x_2}{21}=\frac{3x_2+6}{21},\quad x_2=1,2$$

$$f\left(x_1|x_2\right)=\frac{f\left(x_1,x_2\right)}{f_2\left(x_2\right)}=\frac{\frac{x_1+x_2}{21}}{\frac{3x_2+6}{21}}=\frac{x_1+x_2}{3x_2+6},\quad x_1=1,2,3\quad\text{when}\quad x_2=1,2$$

$$f\left(x_2|x_1\right)=\frac{f\left(x_1,x_2\right)}{f_1\left(x_1\right)}=\frac{\frac{x_1+x_2}{21}}{\frac{2x_1+3}{21}}=\frac{x_1+x_2}{2x_1+3},\quad x_2=1,2\quad\text{when}\quad x_1=1,2,3$$

$$P\left(X_1=2|X_2=2\right)=f(x_1=2|x_2=2)=\frac{4}{12}=\frac{1}{3}$$

$$E\left(X_2|x_1\right)=\sum_{x_2}x_2f\left(x_2|x_1\right)=\sum_{x_2}\frac{x_2(x_1+x_2)}{2x_1+3},\quad x_2=1,2\quad\text{when}\quad x_1=1,2,3$$

$$E\left(X_2|x_1=3\right)=\sum_{x_2}\frac{x_2(3+x_2)}{9}=\frac{1(3+1)}{9}+\frac{2(3+2)}{9}=\frac{14}{9}$$

$$E\left(X_2^2|x_1\right)=\sum_{x_2}x_2^2f\left(x_2|x_1\right)=\sum_{x_2}\frac{x_2^2(x_1+x_2)}{2x_1+3},\quad x_2=1,2\quad\text{when}\quad x_1=1,2,3$$

$$E\left(X_2^2|x_1=3\right)=\sum_{x_2}\frac{x_2^2(3+x_2)}{9}=\frac{1(3+1)}{9}+\frac{4(3+2)}{9}=\frac{24}{9}$$

$$Var\left(X_2|x_1\right)=E\left(X_2^2|x_1\right)-\left\{E\left(X_2|x_1\right)\right\}^2,\ x_2=1,2;\ x_1=1,2,3$$

$$Var\left(X_2|x_1=3\right)=E\left(X_2^2|x_1=3\right)-\left\{E\left(X_2|x_1=3\right)\right\}^2=\frac{24}{9}-\left(\frac{14}{9}\right)^2=\frac{20}{81}$$

□

One note: Check $E\{E(X_2|x_1)\} \stackrel{?}{=} E(X_2)$

$$E\{E(X_2|x_1)\} = E\left\{\sum_{x_2} \frac{x_2(x_1+x_2)}{2x_1+3}\right\} = E\left(\frac{3x_1+5}{2x_1+3}\right) = \sum_{x_1}\left\{\left(\frac{3x_1+5}{2x_1+3}\right)\left(\frac{2x_1+3}{21}\right)\right\} = \frac{33}{21}$$
$$E(X_2) = \sum_{x_2}\left(x_2 \cdot \frac{3x_2+6}{21}\right) = \frac{33}{21}$$

Definition 13. *The conditional expectation of a function of random variables X_1, X_2:*

$$E\{u(X_1,X_2)|x_1\} = \begin{cases} \int_{-\infty}^{\infty} u(x_1,x_2) f(x_2|x_1) dx_2 \\ \sum_{x_2} u(x_1,x_2) f(x_2|x_1) \end{cases}$$

Properties 13. *Conditional expectation:*

1. $E(aX_2+b|x_1) = aE(X_2|x_1) + b$
2. $E(X_2+X_3|x_1) = E(X_2|x_1) + E(X_3|x_1)$
3. $X_2 \geq 0 \quad \Rightarrow \quad E(X_2|x_1) \geq 0$
4. $E\{E(X_2|x_1)\} = E(X_2)$
5. $E\{g(x_1)X_2|x_1\} = g(x_1)E(X_2|x_1)$

Proof.

$$E(aX_2+b|x_1) = \int_{-\infty}^{\infty}(ax_2+b)f(x_2|x_1)dx_2 = a\int_{-\infty}^{\infty} x_2 f(x_2|x_1)dx_2 + b\int_{-\infty}^{\infty} f(x_2|x_1)dx_2$$
$$= aE(X_2|x_1) + b$$

$$E(X_2+X_3|x_1) = E(X_2|x_1) + E(X_3|x_1) \quad \text{(Property of integral)}$$
$$E(X_2|x_1) = \int_{-\infty}^{\infty} x_2 f(x_2|x_1)dx_2 \geq 0 \quad \text{(Definition)}$$

$$E\{E(X_2|x_1)\} = \int_{-\infty}^{\infty} E(X_2|x_1)\, f(x_1)dx_1$$
$$= \int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} x_2 f(x_2|x_1)dx_2\right\} f_1(x_1)dx_1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_2 f(x_2|x_1) f_1(x_1)dx_2 dx_1$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_2 f(x_1,x_2)dx_2 dx_1 \quad \text{(Def of conditional pdf)} = E(X_2)$$

$$E\{g(x_1)X_2|x_1\} = \int_{-\infty}^{\infty} g(x_1)x_2 f(x_2|x_1)dx_2$$
$$= g(x_1)\int_{-\infty}^{\infty} x_2 f(x_2|x_1)dx_2$$
$$= g(x_1)E(X_2|x_1)$$