ST 501: Fundamentals of Statistical Inference

Multivariate probability distributions (Part II)

Department of Statistics, North Carolina State University.

Fall 2024

Covariance and correlation

We now introduce the notion of covariance and correlation between two random variables. These are some of the simplest and yet most widely used (and mis-used) notion of dependence between random variables.

Definition

Let X_1 and X_2 be random variables with $\mathbb{E}[X_1] = \mu_1$ and $\mathbb{E}[X_2] = \mu_2$. The *covariance* between X_1 and X_2 is

$$\mathrm{Cov}(X_1, X_2) = \mathbb{E}[(X_1 - \mu_1)(X_2 - \mu_2)] = \mathbb{E}[X_1 X_2] - \mu_1 \mu_2$$

The correlation between X_1 and X_2 is

$$\mathrm{Cor}(X_1, X_2) = \frac{\mathrm{Cov}(X_1, X_2)}{\sqrt{\mathrm{Var}[X_1]\mathrm{Var}[X_2]}} = \frac{\mathbb{E}[(X_1 - \mu_1)(X_2 - \mu_2)]}{\sqrt{\mathbb{E}[(X_1 - \mu_1)^2]\mathbb{E}[(X_2 - \mu_2)^2]}}$$

We first note some simple properties of covariance.

1. $\mathrm{Cov}(X_1, X_1) = \mathrm{Var}[X_1]$.
2. Covariance is additive, i.e., for any X_1, X_2 and X_3,

$$\begin{aligned}\mathrm{Cov}(X_1 + X_2, X_3) &= \mathbb{E}[(X_1 + X_2)X_3] - (\mu_1 + \mu_2)\mu_3 \\ &= \mathbb{E}[X_1X_3] - \mu_1\mu_3 + \mathbb{E}[X_2X_3] - \mu_2\mu_3 \\ &= \mathrm{Cov}(X_1, X_3) + \mathrm{Cov}(X_2, X_3)\end{aligned}$$

3. $\mathrm{Var}[X_1 + X_2] = \mathrm{Var}(X_1) + 2\mathrm{Cov}(X_1, X_2) + \mathrm{Var}(X_2)$.
4. If X_1 and X_2 are independent, then $\mathrm{Cov}(X_1, X_2) = 0$, i.e.,

$$\mathrm{Var}[X_1 + X_2] = \mathrm{Var}[X_1] + \mathrm{Var}[X_2]$$

   whenever X_1 and X_2 are independent.

Example Let (X, Y) have joint pdf

$$f(x,y) = \begin{cases} 2 & x, y \geq 0, x + y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

We then have

$$\begin{aligned}
\mathbb{E}[X] = \int_0^1 \int_0^1 x f(x,y)\, \mathrm{d}x\mathrm{d}y = \int_0^1 \int_0^{1-x} 2x \mathrm{d}y\mathrm{d}x = \frac{1}{3}, \\
\mathbb{E}[X^2] = \int_0^1 \int_0^{1-x} 2x^2\, \mathrm{d}y\mathrm{d}x = \frac{1}{6}, \\
\mathbb{E}[XY] = \int_0^1 \int_0^{1-x} 2xy\, \mathrm{d}y\mathrm{d}x = \frac{1}{12}, \\
\mathrm{Var}[X] = \mathrm{Var}[Y] = \frac{1}{6} - \left(\frac{1}{3}\right)^2 = \frac{1}{18}, \\
\mathrm{Cov}(X,Y) = \frac{1}{12} - \frac{1}{3} \times \frac{1}{3} = -\frac{1}{36}; \quad \mathrm{Cor}(X,Y) = -\frac{1}{2}.
\end{aligned}$$

Cauchy-Schwarz inequality

Proposition

For any random variable X_1 and X_2 (whose variances exist),

$$\left|\mathbb{E}[X_1 X_2]\right| \leq \sqrt{\mathbb{E}[X_1^2]\mathbb{E}[X_2^2]}.$$

Proof Let $t \in \mathbb{R}$ be arbitrary and consider

$$\mathbb{E}[(X_1 - tX_2)^2] = \mathbb{E}[X_1^2] - 2t\mathbb{E}[X_1X_2] + t^2\mathbb{E}[X_2^2] \geq 0$$

Viewing $f(t) = \mathbb{E}[(X_1 - tX_2)^2]$ as a function of t, the minimum of $f(t)$ occurs when $t = t^*$ where

$$2\mathbb{E}[X_1X_2] + 2t^*\mathbb{E}[X_2^2] = 0 \Longrightarrow t^* = \frac{\mathbb{E}[X_1X_2]}{\mathbb{E}[X_2^2]}$$

$$f(t^*) = \mathbb{E}[X_1^2] - (\mathbb{E}[X_1X_2])^2/\mathbb{E}[X_2^2] \geq 0 \Longrightarrow (\mathbb{E}[X_1X_2])^2 \leq \mathbb{E}[X_1^2]\mathbb{E}[X_2]^2$$

Since the correlation between X_1 and X_2 is defined as

$$\mathrm{Cor}(X_1, X_2) = \frac{\mathbb{E}[(X_1 - \mu_1)(X_2 - \mu_2)]}{\sqrt{\mathbb{E}[(X_1 - \mu_1)^2]\mathbb{E}[(X_2 - \mu_2)^2]}}$$

by taking $\tilde{X}_1 = X_1 - \mu_1$ and $\tilde{X}_2 = X_2 - \mu_2$, we have, by the CS inequality,

$$|\mathrm{Cor}(X_1, X_2)| = \left| \frac{\mathbb{E}[\tilde{X}_1 \tilde{X}_2]}{\sqrt{\mathbb{E}[\tilde{X}_1{}^2]\mathbb{E}[\tilde{X}_2^2]}} \right| \leq 1.$$

Important Equality in the CS inequality holds if and only if

$$\mathbb{E}[(X_1 - t^* X_2)^2] = 0$$

which implies $X_1 = t^* X_2$ everywhere. Thus $|\mathrm{Cor}(X_1, X_2)| = 1$ if and only if there exists a constant c such that $X_1 = cX_2$.

Important Correlation is a measure of **linear dependence** between two random variables.

1. $\mathrm{Cor}(X_1, X_2) = 1$ implies $X_1 = cX_2$ for constant $c > 0$.
2. $\mathrm{Cor}(X_1, X_2) = -1$ implies $X_1 = cX_2$ for constant $c < 0$.
3. X_1 and X_2 are independent implies $\mathrm{Cor}(X_1, X_2) = 0$.
4. Hence, $\mathrm{Cor}(X_1, X_2) = 0$ suggests that X_1 and X_2 are possibly independent.
5. Two random variables are uncorrelated if $\mathrm{Cor}(X_1, X_2) = 0$.
6. If X_1 and X_2 are uncorrelated then

$$\mathrm{Var}[X_1 + X_2] = \mathrm{Var}[X_1] + \mathrm{Var}[X_2]$$

EXAMPLE 5.18 Let X and Y be discrete rv's with joint pmf

$$p(x, y) = \begin{cases} .25 & (x, y) = (-4, 1), (4, -1), (2, 2), (-2, -2) \\ 0 & \text{otherwise} \end{cases}$$

The points that receive positive probability mass are identified on the (x, y) coordinate system in Figure 5.5. It is evident from the figure that the value of X is completely determined by the value of Y and vice versa, so the two variables are completely dependent. However, by symmetry $\mu_X = \mu_Y = 0$ and $E(XY) = (-4)(.25) + (-4)(.25) + (4)(.25) + (4)(.25) = 0$. The covariance is then $\text{Cov}(X,Y) = E(XY) - \mu_X \cdot \mu_Y = 0$ and thus $\rho_{X,Y} = 0$. Although there is perfect dependence, there is also complete absence of any linear relationship!

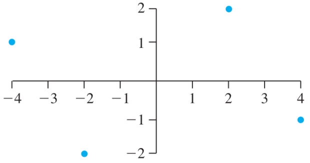

Figure 5.5 The population of pairs for Example 5.18 ■

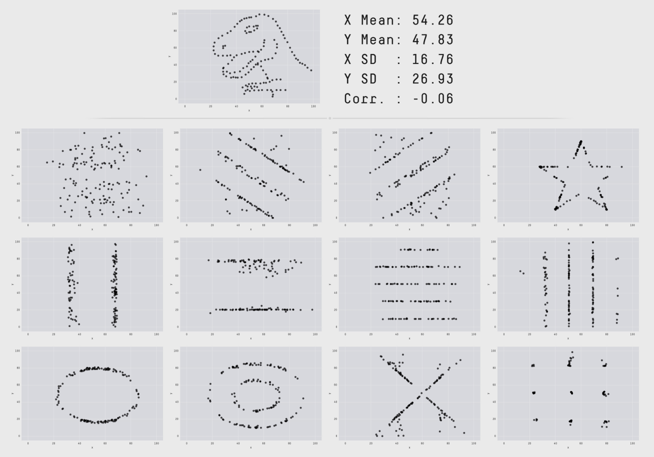

Figure: Dinosaurus

Variance of linear combinations of rvs

We now state a simple yet very useful result for the variance of a linear combination of random variables.

Proposition

Let $X_1, \ldots, X_m$ be random variables and let $a_1, \ldots, a_m$ be constants. Let $U = a_1 X_1 + \cdots + a_m X_m$. Then

$$\mathrm{Var}[U] = \mathrm{Var}[\sum_i a_i X_i] = \sum_i a_i^2 \mathrm{Var}[X_i] + 2 \sum_{i<j} a_i a_j \mathrm{Cov}(X_i, X_j)$$

When the X_i are **pairwise uncorrelated** *then*

$$\mathrm{Var}[U] = \sum_i a_i^2 \mathrm{Var}[X_i].$$

Proof By linearity of the covariance, we have

$$
\begin{aligned}
\mathrm{Var}[U] = \mathrm{Cov}(U,U) &= \mathrm{Cov}\Big(\sum_i a_i X_i, \sum_j a_j X_j\Big) \\
&= \sum_i \mathrm{Cov}\Big(a_i X_i, \sum_j a_j X_j\Big) \\
&= \sum_i \sum_j \mathrm{Cov}(a_i X_i, a_j X_j) \\
&= \sum_i a_i^2 \mathrm{Cov}(X_i, X_i) + \sum_{i \neq j} a_i a_j \mathrm{Cov}(X_i, X_j) \\
&= \sum_i a_i^2 \mathrm{Var}[X_i] + 2 \sum_{i<j} a_i a_j \mathrm{Cov}(X_i, X_j).
\end{aligned}
$$

Similarly, if $U = \sum_i a_i X_i$ and $V = \sum_j b_j Y_j$ then

$$
\mathrm{Cov}(U,V) = \mathrm{Cov}(\sum_i a_i X_i, \sum_j b_j Y_j) = \sum_i \sum_j a_i b_j \mathrm{Cov}(X_i, Y_j)
$$

Variance of a hypergeometric rv

Let $X \sim \mathrm{Hyper}(n, M, N)$ be a hypergeometric random variable. We showed earlier that

$$\mathrm{Var}[X] = \frac{nM(N-M)(N-n)}{N^2(N-1)}$$

We rederive this result by writing X as a sum of indicator rvs.

Write $X = I_1 + I_2 + \cdots + I_n$ where I_j are indicator random variables such that $I_j = 1$ if the jth draw is a "success". Then

$$\mathrm{Var}[X] = \sum_i \mathrm{Var}[I_i] + \sum_{i \neq j} \mathrm{Cov}(I_i, I_j)$$

Some straightforward calculations yield

$$\mathbb{E}[I_i] = \frac{M}{N}, \quad \mathbb{E}[I_i I_j] = \frac{M(M-1)}{N(N-1)},$$
$$\mathrm{Cov}(I_i, I_j) = \frac{M(M-1)}{N(N-1)} - \frac{M^2}{N^2} = -\frac{M(N-M)}{N^2(N-1)},$$
$$\mathrm{Var}[I_i] = \frac{M}{N} \times \left(1 - \frac{M}{N}\right) = \frac{M(N-M)}{N^2}.$$

We therefore have

$$\begin{aligned}
\mathrm{Var}[X] &= \sum_i \mathrm{Var}[I_i] + \sum_{i \neq j} \mathrm{Cov}(I_i, I_j) \\
&= \frac{nM(N-M)}{N^2} - \frac{n(n-1)M(N-M)}{N^2(N-1)}. \\
&= \frac{nM(N-M)(N-n)}{N^2(N-1)}
\end{aligned}$$

Variance of the sample mean

Proposition

Let $X_1, \ldots, X_n$ *be independent and identically distributed rvs. Denote* $\mathbb{E}[X_i] \equiv \mu$ *and* $\mathrm{Var}[X_i] \equiv \sigma^2$*. Let* $\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ *be the sample mean of the* $\{X_i\}$*'s. Then*

$$\mathrm{Var}[\bar{X}] = \mathrm{Var}\left[\frac{1}{n}\sum_i X_i\right] = \frac{1}{n^2}\mathrm{Var}[\sum_i X_i] = \frac{1}{n^2}\sum_i \mathrm{Var}[X_i] = \frac{\sigma^2}{n}.$$

Important As $\mathbb{E}[\bar{X}] = \mu$, we have by Chebyshev's inequality that

$$P\left(|\bar{X} - \mu| \geq \frac{k\sigma}{\sqrt{n}}\right) \leq \frac{1}{k^2}$$

Thus, for any any fixed $\epsilon > 0$,

$$P(|\bar{X} - \mu| \geq \epsilon) \longrightarrow 0$$

as $n \to \infty$. In summary, $\bar{X}$ is a consistent estimate of μ.

See Problem 6.31 in your textbook for more details.

Multinomial distribution

We now introduce one of the most famous multivariate discrete distribution. The multinomial distribution generalizes the binomial distribution to $K \geq 3$ outcomes.

Definition

A multinomial experiment is an experiment in which

1. There are n independent and identical trials (n fixed).
2. Each trial results in exactly 1 out of K different outcomes.
3. The ith trial result in the kth outcome with probability p_k.

Associated to a multinomial experiment is a multivariate rv $X = (X_1, X_2, \ldots, X_K)$ with X_k being the number of trials for which the kth outcome occurs. Note that $\sum_k X_k = n$.

Definition
Let X be a multinomial rv with n trials and K outcome. Let $(p_1, \ldots, p_K)$ be the probability of the outcomes. Then X has joint pmf

$$p(x_1, x_2, \ldots, x_K) = \frac{n!}{x_1!x_2! \ldots x_K!} p_1^{x_1} p_2^{x_2} \ldots p_K^{x_k}$$

Example 5.23. The State Hygienic Laboratory at the University of Iowa tests thousands of Iowa residents each year for chlamydia (CT) and gonorrhea (NG). On a given day, suppose the lab receives $n = 100$ specimens to be tested. Define

Category 1:	CT−/NG−	$(p_1 = 0.90)$
Category 2:	CT+/NG−	$(p_2 = 0.07)$
Category 3:	CT−/NG+	$(p_3 = 0.02)$
Category 4:	CT+/NG+	$(p_4 = 0.01)$

and let $\mathbf{Y} = (Y_1, Y_2, Y_3, Y_4)$ denote the category counts observed after testing. Envisioning each specimen as a "trial," regarding the specimens as mutually independent, and assuming the category probabilities in $\mathbf{p} = (p_1, p_2, p_3, p_4)$ are the same for each specimen, then

$$\mathbf{Y} \sim \text{mult}\left(n = 100, \mathbf{p}; \sum_{j=1}^{4} p_j = 1\right).$$

Example 5.23. The State Hygienic Laboratory at the University of Iowa tests thousands of Iowa residents each year for chlamydia (CT) and gonorrhea (NG). On a given day, suppose the lab receives $n = 100$ specimens to be tested. Define

Category 1:	CT−/NG−	$(p_1 = 0.90)$
Category 2:	CT+/NG−	$(p_2 = 0.07)$
Category 3:	CT−/NG+	$(p_3 = 0.02)$
Category 4:	CT+/NG+	$(p_4 = 0.01)$

and let $\mathbf{Y} = (Y_1, Y_2, Y_3, Y_4)$ denote the category counts observed after testing. Envisioning each specimen as a "trial," regarding the specimens as mutually independent, and assuming the category probabilities in $\mathbf{p} = (p_1, p_2, p_3, p_4)$ are the same for each specimen, then

$$\mathbf{Y} \sim \text{mult}\left(n = 100, \mathbf{p}; \sum_{j=1}^{4} p_j = 1\right).$$

Suppose we are interested in the probability that 88 specimens are category 1, 10 are category 2 and category 3 and 4 both has 1 specimen. Then

$$P(Y_1 = 88, \ldots, Y_4 = 1) = \frac{100!}{88!10!1!1!}(0.9)^{88}(0.07)^{10}(0.02)^{1}(0.01)^{1}.$$

Example: From Poisson to multinomial

Q. Let $X_1, X_2, \ldots, X_K$ be independent Poisson random variables with rate parameters $\lambda_1, \lambda_2, \ldots, \lambda_K$. Let $S = X_1 + X_2 + \cdots + X_K$. What is the distribution of $X = (X_1, X_2, \ldots, X_K)$ given S ?

Example: From Poisson to multinomial

Q. Let $X_1, X_2, \ldots, X_K$ be independent Poisson random variables with rate parameters $\lambda_1, \lambda_2, \ldots, \lambda_K$. Let $S = X_1 + X_2 + \cdots + X_K$. What is the distribution of $X = (X_1, X_2, \ldots, X_K)$ given S ?

Let $\lambda = \lambda_1 + \cdots + \lambda_K$. The joint pmf of X given $S = n$ is

$$p(x_1, \ldots, x_k \mid S = n) = \frac{(e^{-\lambda_1}\lambda_1^{x_1}/x_1!)(e^{-\lambda_2}\lambda_2^{x_2}/x_2!)\ldots(e^{\lambda_K}\lambda_K^{x_K}/x_K!)}{e^{-\lambda}\lambda^n/n!}$$

$$= \frac{n!}{x_1!x_2!\ldots x_K!} \times \left(\frac{\lambda_1}{\lambda}\right)^{x_1}\left(\frac{\lambda_2}{\lambda}\right)^{x_2}\ldots\left(\frac{\lambda_K}{\lambda}\right)^{x_K}$$

which is the pmf of a multinomial with $p = (\frac{\lambda_1}{\lambda}, \ldots, \frac{\lambda_K}{\lambda})$.

88. Each customer making a particular Internet purchase must pay with one of three types of credit cards (think Visa, MasterCard, AmEx). Let A_i ($i = 1, 2, 3$) be the event that a type i credit card is used, with $P(A_1) = .5$, $P(A_2) = .3$, and $P(A_3) = .2$. Suppose that the number of customers who make such a purchase on a given day is a Poisson rv with parameter λ. Define rv's X_1, X_2, X_3 by X_i = the number among the N customers who use a type i card ($i = 1, 2, 3$). Show that these three rv's are independent with Poisson distributions having parameters $.5\lambda$, $.3\lambda$, and $.2\lambda$, respectively. [*Hint:* For non-negative integers x_1, x_2, x_3, let $n = x_1 + x_2 + x_3$. Then $P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = P(X_1 = x_1, X_2 = x_2, X_3 = x_3, N = n)$ [why is this?]. Now condition on $N = n$, in which case the three Xi's have a trinomial distribution (multinomial with three categories) with category probabilities .5, .3, and .2.]

Proposition

Let $X = (X_1, X_2, \ldots, X_K)$ *be a multinomial rv with* n *trials and outcome probabilities* $(p_1, \ldots, p_K)$*. Then*

$$\mathbb{E}[X_k] = np_k,\ \mathrm{Var}[X_k] = np_k(1 - p_k)$$
$$\mathrm{Cov}(X_k, X_\ell) = -np_k p_\ell \text{ if } k \neq \ell$$

Proof The formulas for $\mathbb{E}[X_k]$ and $\mathrm{Var}[X_k]$ follows directly from the formulas for a binomial rv with n trials and success probability p_k

We now derive the formula for $\mathrm{Cov}(X_k, X_\ell)$.

Write $X_k = I_1^{(k)} + I_2^{(k)} + \cdots + I_n^{(k)}$ where $I_i^{(k)}$ is the indicator random variable with $I_i^{(k)} = 1$ if the ith trial result in the kth outcome and $I_i^{(k)} = 0$ otherwise. We have

$$\mathrm{Cov}(X_k, X_\ell) = \mathrm{Cov}\Big(\sum_i I_i^{(k)}, \sum_j I_j^{(\ell)}\Big) = \sum_i \sum_j \mathrm{Cov}(I_i^{(k)}, I_j^{(\ell)})$$

Straightforward calculations yield (for $k \neq \ell$),

$$\mathbb{E}[I_i^{(k)}] = p_k \quad \text{for all } i = 1, 2, \ldots, n$$
$$\mathbb{E}[I_i^{(k)} I_i^{(\ell)}] = 0 \text{ for all } i = 1, 2, \ldots, n,$$
$$\mathbb{E}[I_i^{(k)} I_j^{(\ell)}] = \mathbb{E}[I_i^{(k)}]\mathbb{E}[I_j^{(\ell)}] \text{ if } i \neq j$$

We therefore have

$$\mathrm{Cov}(I_i^{(k)}, I_i^{(\ell)}) = -p_k p_\ell,$$
$$\mathrm{Cov}(I_i^{(k)}, I_j^{(\ell)}) = 0, \quad \text{for all } i \neq j,$$
$$\mathrm{Cov}(X_k, X_\ell) = \sum_i \mathrm{Cov}(I_i^{(k)}, I_i^{(\ell)}) = -np_k p_\ell.$$

Conditional Expectation

We first define conditional expectation for discrete bivariate r.v.

Definition
Let $X = (X_1, X_2)$ be a discrete bivariate r.v. For any function g the conditional expectation of $g(X_1)$ given $X_2 = x_2$ is

$$\mathbb{E}[g(X_1) \mid X_2 = x_2] = \sum_{x_1} g(x_1) p(x_1 \mid x_2)$$

where the summation is over all x_1 with $p(x_1, x_2) > 0$. Here we have implicitly assumed that $P(X_2 = x_2) > 0$.

Important We note that $\mathbb{E}[g(X_1) \mid X_2 = x]$ is a function of x_2.

Furthermore, since $p(\cdot \mid x_2)$ is a valid pmf, $\mathbb{E}[g(X_1) \mid X_2 = x_2]$ is the expected value of $g(X_1)$ under the conditional pmf $p(\cdot \mid x_2)$.

Important If we do not condition on the value of X_2 then we can define

$$\mathbb{E}[g(X_1) \mid X_2] = \sum_{x_1} g(x_1)p(x_1 \mid X_2)$$

which is now a random variable.

In other words $\mathbb{E}[g(X_1) \mid X_2] = h(X_2)$ for some function h.

If X_2 is independent of X_1 then $\mathbb{E}[g(X_1) \mid X_2] = \mathbb{E}[g(X_1)]$.

Important The above definition also implies, for any set A,

$$\begin{aligned}\mathbb{E}[g(X_1)I(X_2 \in A)] &= \sum_{x_1} \sum_{x_2 \in A} g(x_1)p(x_1, x_2) \\ &= \sum_{x_2 \in A} \sum_{x_1} g(x_1)p(x_1 \mid x_2)p_2(x_2) \\ &= \mathbb{E}[\mathbb{E}[g(X_1) \mid X_2]I(X_2 \in A)]\end{aligned}$$

where the (unconditional) expectation is taken wrt X_2.

Important Taking $A = \mathbb{R}$, $I(X_2 \in A) \equiv 1$ and hence

$$\mathbb{E}[g(X_1)] = \mathbb{E}[\mathbb{E}[g(X_1) \mid X_2]I(X_2 \in \mathbb{R})] = \mathbb{E}[\mathbb{E}[g(X_1) \mid X_2]]$$

This is known as the tower property of conditional expectation.

We now define the notion of conditional expectation for continuous bivariate random variables.

Definition Let $X = (X_1, X_2)$ be a continuous bivariate r.v. Then for any function g, the conditional expectation of $g(X_1)$ given $X_2 = x_2$ is defined as

$$\mathbb{E}[g(X_1) \mid X_2 = x_2] = \int_{-\infty}^{\infty} g(x_1) f(x_1 \mid x_2) \, \mathrm{d}x_1$$

Important $\mathbb{E}[g(X_1) \mid X_2 = x_2]$ is a function of x_2 such that, for all sets A

$$\begin{aligned}\mathbb{E}[g(X_1)I(X_2 \in A)] &= \int_A \int_{-\infty}^{\infty} g(x_1) f(x_1, x_2)\,\mathrm{d}x_1 \mathrm{d}x_2 \\ &= \int_A \mathbb{E}[g(X_1) \mid X_2 = x_2] f_2(x_2)\,\mathrm{d}x_2\end{aligned}$$

Letting $A = \mathbb{R}$, we once again have

$$\mathbb{E}[g(X_1)] = \mathbb{E}[\mathbb{E}[g(X_1) \mid X_2]].$$

Furthermore, letting $A = [x_2 - \epsilon, x_2 + \epsilon]$ for some "small" $\epsilon > 0$

$$\begin{aligned}\mathbb{E}[g(X_1) \mid X_2 = x_2] &\approx \mathbb{E}[g(X_1) \mid X_2 \in (x_2 - \epsilon, x_2 + \epsilon)] \\ &\approx \frac{\mathbb{E}[g(X_1)I(X_2 \in (x_2 - \epsilon, x_2 + \epsilon))]}{P(X_2 \in (x_2 - \epsilon, x_2 + \epsilon))}\end{aligned}$$

Example Let $Y = (Y_1, Y_2)$ have joint pdf

$$f(y_1, y_2) = 1/2 \quad 0 \leq y_1 \leq y_2 \leq 2.$$

Find $\mathbb{E}[Y_1 \mid Y_2 = y_2]$.

We need to find $f(y_1 \mid y_2) = f(y_1, y_2)/f_2(y_2)$. The marginal pdf for Y_2 is

$$\int_0^2 f(y_1, y_2)\,\mathrm{d}y_1 = \int_0^{y_2} 1/2\,\mathrm{d}y_2 = y_2/2$$

and hence $f(y_1, \mid y_2) = 1/y_2$ for $y_1 \leq y_2$.

We therefore have

$$\mathbb{E}[Y_1 \mid Y_2 = y_2] = \int_0^2 y_1 f(y_1 \mid y_2) = \int_0^{y_2} y_1/y_2\,\mathrm{d}y_1 = \frac{1}{2y_2} y_1^2 \Big|_0^{y_2} = y_2/2.$$

EXAMPLE **5.32** A quality control plan for an assembly line involves sampling $n = 10$ finished items per day and counting Y, the number of defectives. If p denotes the probability of observing a defective, then Y has a binomial distribution, assuming that a large number of items are produced by the line. But p varies from day to day and is assumed to have a uniform distribution on the interval from 0 to $1/4$. Find the expected value of Y.

This problem doesn't fit into the two definitions above as the pair (Y, p) is neither bivariate continuous nor bivariate discrete but rather a mixed of discrete and continuous. There is thus neither a conditional pdf nor a conditional pmf for $Y \mid p$.

Nevertheless, we can interpret the problem as saying that the marginal pmf for Y is

$$P(Y = k) = \int_0^{1/4} \binom{n}{k} p^k (1-p)^{n-k} \, \mathrm{d}p$$

and hence

$$\begin{aligned}
\mathbb{E}[Y] &= \sum_{k=0}^{n} k \int_0^{1/4} \binom{n}{k} p^k (1-p)^{n-k} \, \mathrm{d}p \\
&= \int_0^{1/4} \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \, \mathrm{d}p \\
&= \int_0^{1/4} np \, \mathrm{d}p = \frac{n}{8}
\end{aligned}$$

We note that the integral $\int_0^{1/4} np \, \mathrm{d}p$ behaves like $\mathbb{E}[\mathbb{E}[Y \mid p]]$.

Bivariate normal distribution

Q. Let X_1 and X_2 be standard normals. Define

$$U_1 = \sigma_1 X_1 + \mu_1, \quad U_2 = \sigma_2(\rho X_1 + \sqrt{1-\rho^2} X_2) + \mu_2$$

Here $\rho \in [-1, 1]$, and both σ_1 and σ_2 are positive numbers. What is the joint pdf for $U = (U_1, U_2)$?

A. We have

$$\begin{aligned}(U_2 \mid U_1 = u_1) &= (U_2 \mid X_1 = (u_1 - \mu_1)/\sigma_1) \\ &= \frac{\sigma_2 \rho (u_1 - \mu_1)}{\sigma_1} + \sigma_2 \sqrt{1-\rho^2} X_2 + \mu_2\end{aligned}$$

and since X_2 is standard normal,

$$(U_2 \mid U_1 = u_1) \sim \mathcal{N}(\mu_2 + \frac{\sigma_2 \rho}{\sigma_1}(u_1 - \mu_1), (1-\rho)^2 \sigma^2)$$

We can reverse the role of U_1 and U_2, in which case

$$(U_1 \mid U_2 = u_2) \sim \mathcal{N}(\mu_1 + \frac{\sigma_1 \rho}{\sigma_2}(u_2 - \mu_2), (1 - \rho^2)\sigma_1^2).$$

The joint pdf for $U = (U_1, U_2)$ is then

$$\begin{aligned} f_U(u) &= f_{U_1|U_2}(u_1 \mid u_2) \times f_{U_2}(u_2) \\ &= \tfrac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{1/2}} \exp\Big(\tfrac{-1}{1-\rho^2}\Big(\tfrac{(u_1-\mu_1)^2}{2\sigma_1^2} - \tfrac{\rho(u_1-\mu_1)(u_2-\mu_2)}{\sigma_1\sigma_2} + \tfrac{(u_2-\mu_2)^2}{2\sigma_2^2}\Big)\Big), \end{aligned}$$

quite a montrosity.

Law of total variation

Definition Let X_1 and X_2 be two random variables. Define the conditional variance of X_1 given X_2 as

$$\mathrm{Var}[X_1 \mid X_2] = \mathbb{E}[X_1^2 \mid X_2] - (\mathbb{E}[X_1 \mid X_2])^2$$

We emphasize that $\mathrm{Var}[X_1 \mid X_2]$ is a **random variable**.

Proposition For any two random variables X_1 and X_2,

$$\mathrm{Var}[X_1] = \mathbb{E}[\mathrm{Var}[X_1 \mid X_2]] + \mathrm{Var}[\mathbb{E}[X_1 \mid X_2]$$

Proof From the definition of $\mathrm{Var}[X_1 \mid X_2]$ we have

$$\begin{aligned}\mathbb{E}[\mathrm{Var}[X_1 \mid X_2]] &= \mathbb{E}[\mathbb{E}[X_1^2 \mid X_2]] - \mathbb{E}[(\mathbb{E}[X_1 \mid X_2])^2] \\ &= \mathbb{E}[X_1^2] - \mathbb{E}[(\mathbb{E}[X_1 \mid X_2])^2]\end{aligned}$$

Furthermore, by the tower property of conditional expectation,

$$\begin{aligned}\mathrm{Var}[\mathbb{E}[X_1 \mid X_2]] &= \mathbb{E}[(\mathbb{E}[X_1 \mid X_2])^2] - (\mathbb{E}[\mathbb{E}[X_1 \mid X_2]])^2 \\ &= \mathbb{E}[(\mathbb{E}[X_1 \mid X_2])^2] - (\mathbb{E}[X_1])^2\end{aligned}$$

Adding the two terms yield the claim.

Example A hen lays X eggs where X is Poisson with rate parameter λ. Each egghatches with probablity p, independently of the other, yielding Y chicks. Find $\mathbb{E}[Y]$ and $\mathrm{Var}[Y]$.

Using the tower property, we have

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid X]] = \mathbb{E}[Xp] = \lambda p.$$

Using the law of total variation we have

$$\begin{aligned}\mathrm{Var}[Y] &= \mathbb{E}[\mathrm{Var}[Y \mid X]] + \mathrm{Var}[\mathbb{E}[Y \mid X] \\ &= \mathbb{E}[Xp(1-p)] + \mathrm{Var}[Xp] = \lambda p(1-p) + \lambda p^2 = \lambda p\end{aligned}$$

As $\mathbb{E}[Y] = \mathrm{Var}[Y] = \lambda p$, this suggests that $Y \sim \mathrm{Pois}(\lambda p)$ which we have seen earlier.

Example Let N be a random variable taking positive integer values and let $S = X_1 + X_2 + \cdots + X_N$ where the X_i are independent and identically distributed random variables. Suppose also that N is independent of the X_i. Find $\mathbb{E}[S]$ and $\mathrm{Var}[S]$.

Once again, by the tower property

$$\mathbb{E}[S] = \mathbb{E}[\mathbb{E}[S \mid N]] = \mathbb{E}[N\mathbb{E}[X_1]] = \mathbb{E}[N] \times \mathbb{E}[X_1]$$

By the law of total variance, we have

$$\begin{aligned}
\mathrm{Var}[S] &= \mathrm{Var}[\mathbb{E}[S \mid N]] + \mathbb{E}[\mathrm{Var}[S \mid N]] \\
&= \mathrm{Var}[N\mathbb{E}[X_1]] + \mathbb{E}[N\mathrm{Var}[X_1]] \\
&= \mathrm{Var}[N] \times (\mathbb{E}[X_1])^2 + \mathbb{E}[N] \times \mathrm{Var}[X_1]
\end{aligned}$$