Chapter 1
Special Distributions

1. **Special Distributions**
   Bernoulli, binomial, geometric, and negative binomial
   Sampling with and without replacement; Hypergeometric
   Finite sample variance correction
   Poisson and an "informal" Poisson process
   Stationary and independent increments
   Exponential and Gamma; Strong Markov property
   Normal, and the classical CLT; Chi-square
   Uniform, beta, uniform order statistics
   Cauchy
   Rademacher, and symmetrization
   Multinomial, and its moments

2. **Convolution and related formulas**
   Sums, products, and quotients
   Student's t; Snedecor's F; and beta

3. **The multivariate normal distribution**
   Properties of covariance matrices
   Characteristic function
   Marginals, independence, and linear combinations
   Linear independence
   The multivariate normal density
   Conditional densities
   Facts about Chi-square distributions

4. **General integration by parts formulas**
   Representations of random variables
   Formulas for means, variances, and covariances via integration by parts

Chapter 1

Special Distributions

1 Special Distributions

Independent Bernoulli Trials

If $P(X=1)=p=1-P(X=0)$, then X is said to be a *Bernoulli*(p) random variable. We refer to the event $[X=1]$ as success, and to $[X=0]$ as failure.

Let $X_1,\ldots,X_n$ be i.i.d. Bernoulli(p), and let $S_n=X_1+\cdots+X_n$ denote the number of successes in n independent Bernoulli(p) trials. Now

$$P(X_i=x_i,\ i=1,\ldots,n)=p^{\sum_1^n x_i}(1-p)^{n-\sum_1^n x_i}$$

if all x_i equal 0 or 1; this formula gives the joint distribution of $X_1,\ldots,X_n$. From this we obtain

(1) $\qquad P(S_n=k)=\binom{n}{k}p^k(1-p)^{n-k} \qquad \text{for } \ k=0,\ldots,n,$

since each of the $\binom{n}{k}$ different placings of k 1's in an $n-$vector containing k 1's and $n-k$ 0's has probability $p^k(1-p)^{n-k}$ from the previous sentence. We say that $S_n\sim$ *Binomial*(n,p) when (1) holds. Note that Binomial$(1,p)$ is the same as Bernoulli(p).

Let $X_1,X_2,\ldots$ be i.i.d. Bernoulli(p). Let $Y_1\equiv W_1\equiv\min\{n:\ S_n=1\}$. Since $[Y_1=k]=[X_1=0,\ldots,X_{k-1}=0,X_k=1]$, we have

(2) $\qquad P(Y_1=k)=(1-p)^{k-1}p \qquad \text{for } \ k=1,2,\ldots\ .$

We say that $Y_1\sim$ *Geometric*(p). Now let $W_m\equiv\min\{n:\ \ S_n=m\}$. We call W_m *the waiting time to the* m-th *success*. Let $Y_m\equiv W_m-W_{m-1}$ for $m\ge 1$, with $W_0\equiv 0$; we call the Y_m's the *interarrival times*. Note that $[W_m=k]=[S_{k-1}=m-1,X_k=1]$. Hence

(3) $\qquad P(W_m=k)=\binom{k-1}{m-1}p^m(1-p)^{k-m} \qquad \text{for } \ k=m,m+1,\ldots\ .$

We say that $W_m\sim$ *Negative Binomial*(m,p).

Exercise 1.1 Show that $Y_1,Y_2,\ldots$ are i.i.d. Geometric(p).

Since the number of successes in n_1+n_2 trials is the number of successes in the first n_1 trials plus the number of successes in the next n_2 trials, it is clear that for independent $Z_i\sim$Binomial(n_i,p),

(4) $\qquad Z_1+Z_2\sim Binomial(n_1+n_2,p).$

Likewise, for independent $Z_i \sim$*Negative Binomial*(m_i, p),

$$(5) \qquad Z_1 + Z_2 \sim \textit{Negative Binomial}(m_1 + m_2, p)\,.$$

Urn Models

Suppose that an urn contains N balls of which M bear the number 1 and $N - M$ bear the number 0. Thoroughly mix the balls in the urn. Draw one ball at random. Let X_1 denote the number on the ball. Then $X_1 \sim$ Bernoulli(p) with $p = M/N$. Now replace the ball back in the urn, thoroughly mix, and draw at random a second ball with number X_2, and so forth. Let $S_n = X_1 + \cdots + X_n \sim$ Binomial(n, p) with $p = M/N$.

Suppose now that the same scheme is repeated except that the balls are not replaced. In this *sampling without replacement* scheme $X_1, \ldots, X_n$ are dependent Bernoulli(p) random variables with $p = M/N$. Also

$$(6) \qquad P(S_n = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

provided the value k is possible (i.e. $k \le M$ and $n - k \le N - M$). We say that $S_n \sim$ *Hypergeometric*(N, M, n).

Suppose now that sampling is done without replacement, but the N balls in the urn bear the numbers $a_1, \ldots, a_N$. Let $X_1, \ldots, X_n$ denote the numbers on the first n balls drawn, and let $S_n \equiv X_1 + \cdots + X_n$. We call this the *finite sampling model*. Call $\overline{a} \equiv \sum_1^N a_i/N$ and $\sigma_a^2 \equiv \sum_1^N (a_i - \overline{a})^2/N$ the *population mean* and *population variance*. Note that X_i has expectation $\overline{a}$ and variance σ_a^2 for all $i = 1, \ldots, n$, since we now assume $n \le N$. Now from the formula for the variance of a sum of random variables and symmetry we have

$$(7) \qquad 0 = Var\left(\sum_1^N X_i\right) = NVar(X_1) + N(N-1)Cov(X_1, X_2)$$

since $\sum_1^N X_i$ is a constant. Thus

$$(8) \qquad Cov[X_1, X_2] = -\sigma_a^2/(N-1)\,.$$

Thus an easy computation gives

$$(9) \qquad Var[S_n/n] = \frac{\sigma_a^2}{n}\left(1 - \frac{n-1}{N-1}\right)\,,$$

where $(1 - (n-1)/(N-1))$ is called the *correction factor for finite sampling*.

Exercise 1.2 Verify (8) and (9).

Exercise 1.3 If $X \sim$ Binomial(m, p) and $Y \sim$ Binomial(n, p) are independent, then the conditional distribution of X given that $X + Y = N$ is Hypergeometric$(m + n, N, m)$.

The Poisson Process

Suppose now that $X_{n1}, X_{n2}, \ldots,$ are i.i.d. Bernoulli(p_n) where $np_n \to \lambda$ as $n \to \infty$. Let $S_n = X_{n1} + \cdots + X_{nn}$ so that $S_n \sim$ Binomial(n, p_n). An easy calculation shows that

$$(10) \qquad P(S_n = k) \to \frac{\lambda^k}{k!}e^{-\lambda} \qquad \text{for} \quad k = 0, 1, \ldots.$$

If $P(S = k) = \lambda^k e^{-\lambda}/k!$ for $k = 0, 1, \ldots$, then we say that $S \sim$ *Poisson*(λ). The above can be used to model the following Geiger counter experiment. A radioactive source with "large" half - life is placed near a Geiger counter. Let $\mathbb{N}(t)$ denote the number of particles registered by time t; we will say that $\{\mathbb{N}(t) : \ t \geq 0\}$ is a *Poisson process*. (Do note that our treatment is purely informal.) Physical considerations lead us to believe that $\mathbb{N}(t_1), \mathbb{N}(t_1, t_2], \cdots, \mathbb{N}(t_{k-1}, t_k]$ should be independent random variables where $\mathbb{N}(t_{i-1}, t_i]$ denotes the *increment* $\mathbb{N}(t_i) - \mathbb{N}(t_{i-1})$; we say that $\mathbb{N}$ has *independent increments*. We now define

(11) $\quad \lambda \equiv E\mathbb{N}(1) = \text{the } \textit{intensity} \text{ of the process}\,.$

Let M denote the number of radioactive particles in our source, and let X_i equal 1 or 0 depending on whether or not the $i-$th particle registers by time $= 1$ or not. It seems a reasonable model to assume that $X_1, \ldots, X_M$ are i.i.d. Bernoulli. Since $\mathbb{N}(1) = X_1 + \cdots + X_M$ has mean $\lambda = E\mathbb{N}(1) = ME(X_1)$, this leads to $\mathbb{N}(1) \sim \text{Binomial}(M, \lambda/M)$. By the first paragraph of this section $\mathbb{N}(1)$ is thus approximately a Poisson(λ) random variable. We now alter our point of view slightly, and *use this approximation as our model*.

Thus $\mathbb{N}(1)$ is a Poisson(λ) random variable. By the stationary and independent increments we thus have

(12) $\quad \mathbb{N}(s, t] \sim \textit{Poisson}(\lambda(t-s)) \qquad \text{for all } \ 0 \leq s \leq t$

while

(13) $\quad \mathbb{N} \quad \text{has independent increments}\,.$

Note also that $\mathbb{N}(0) = 0$. (This is actually enough to rigorously specify a Poisson process.)

Let $Y_1 \equiv W_1 \equiv \inf\{t > 0 : \ \mathbb{N}(t) = 1\}$. Since

(14) $\quad [Y_1 > t] = [\mathbb{N}(t) = 0],$

we see that $P(Y_1 > t) = P(\mathbb{N}(t) = 0) = e^{-\lambda t}$ by (12). Thus Y_1 has distribution function $1 - \exp(-\lambda t)$ for $t \geq 0$ and density

(15) $\quad f_{Y_1}(t) = \lambda e^{-\lambda t} \qquad \text{for} \qquad t \geq 0\,;$

we say that $Y_1 \sim$ *Exponential*(λ). Now let $W_m \equiv \inf\{t > 0 : \ \mathbb{N}(t) = m\}$; we call W_m the $m-$th *waiting time*. We call $Y_m \equiv W_m - W_{m-1}$, $m \geq 1$, the *interarrival times*. In light of the physical properties of our Geiger counter model, and using (13), it seems reasonable that

(16) $\quad Y_1, \, Y_2, \, \ldots \quad \text{are i.i.d. Exponential}(\lambda)\,.$

Our assumption of the previous sentence could be expressed as

(17) $\quad \begin{array}{l} Y_1 \ \text{ and } \ \mathbb{N}_1(t) \equiv \mathbb{N}(Y_1, Y_1 + t] \ \text{ are independent} \\ \text{and } \ \mathbb{N}_1 \ \text{ is again a Poisson process with intensity } \ \lambda\,; \end{array}$

we will call this the *strong Markov property* of the Poisson process. Now

(18) $\quad [W_m > t] = [\mathbb{N}(t) < m]\,,$

so that $P(W_m > t) = \sum_{k=0}^{m-1} (\lambda t)^k e^{-\lambda t}/k!$; differentiating this expression shows that W_m has density

(19) $\quad f_{W_m}(t) = \lambda^m t^{m-1} e^{-\lambda t}/\Gamma(m) \qquad \text{for } \ t \geq 0\,;$

we say that $W_m \sim$ *Gamma*(m, λ). Contained in this is a proof that for independent $Z_i \sim \text{Gamma}(m_i, \lambda)$,

(20) $\quad Z_1 + Z_2 \sim \text{Gamma}(m_1 + m_2, \lambda).$

Exercise 1.4 Verify (10).

Exercise 1.5 Verify (16).

Exercise 1.6 Verify (19).

It is true that (19) is a density for any real number $m > 0$; and the property (20) still holds for real m_i's.

Exercise 1.7 If $X \sim$ Poisson(λ_1) and $Y \sim$ Poisson(λ_2) are independent, then the conditional distribution of X given $X + Y = n$ is *Binomial*$(n, \lambda_1/(\lambda_1 + \lambda_2))$.

Exercise 1.8 If $X \sim$ Gamma(α, λ) and $Y \sim$ Gamma(β, λ) are independent, show that $X/(X + Y) \sim$ Beta(α, β); i.e. $U \equiv X/(X+Y)$ has density $\{\Gamma(\alpha+\beta)/\Gamma(\alpha)\Gamma(\beta)\}u^{\alpha-1}(1-u)^{\beta-1}$, $0 < u < 1$.

The Normal Distribution

Suppose that the random variable Z has density

$$(21) \qquad \phi(z) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{z^2}{2}) \qquad \text{for} \quad -\infty < z < \infty;$$

then Z is said to be a *standard normal random variable*. We let the corresponding distribution function be denoted by Φ. Thus

$$(22) \qquad \Phi(z) = P(Z \le z) = \int_{\infty}^{z} \phi(y)dy\,.$$

If $b > 0$, then $F_{a+bZ}(x) = P(a + bZ \le x) = P(Z \le (x - a)/b) = \Phi((x - a)/b)$. Thus $a + bZ$ has density

$$(23) \qquad f_{a+bZ}(x) = \frac{1}{b}\phi\left(\frac{x-a}{b}\right) \qquad \text{for} \quad -\infty < x < \infty\,.$$

Note that (23) holds for $Z \sim f_Z$ if we replace ϕ by f_Z.

Exercise 1.9 Show that ϕ given in (21) is a density. Show that this density has mean 0 and variance 1.

Thus $X \equiv \mu + \sigma Z \sim (\mu, \sigma^2)$ with density

$$(24) \qquad \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \qquad \text{for} \quad -\infty < x < \infty\,;$$

we say that $X \sim$ *Normal*(μ, σ^2) or just $N(\mu, \sigma^2)$.

The importance of the normal distribution derives from the following theorem. Recall from the properties of expectation and variance that if $X_1, \ldots, X_n$ are i.i.d. (μ, σ^2), then $\sqrt{n}(\overline{X}_n - \mu)/\sigma$ has mean 0 and variance 1 where $\overline{X}_n \equiv (X_1 + \cdots + X_n)/n$. But much more is true.

Theorem 1.1 (Classic CLT). If $X_1, \ldots, X_n$ are i.i.d. (μ, σ^2), then

$$\sqrt{n}(\overline{X}_n - \mu) \to_d N(0, \sigma^2) \qquad \text{as } n \to \infty\,. \tag{25}$$

Hence if $\sigma > 0$

$$\sqrt{n}(\overline{X}_n - \mu)/\sigma \to_d N(0, 1) \qquad \text{as } n \to \infty\,. \tag{26}$$

This result will be stated again in Chapter 2 along with other central limit theorems. We will use it in the meantime for motivational purposes.

Suppose that Z is $N(0, 1)$. Then

$$\begin{aligned} F_{Z^2}(x) &= P(Z^2 \le x) = P(-\sqrt{x} \le Z \le \sqrt{x}) \\ &= F_Z(\sqrt{x}) - F_Z(-\sqrt{x}) \\ &= \Phi(\sqrt{x}) - \Phi(-\sqrt{x})\,; \end{aligned} \tag{27}$$

thus Z^2 has density

$$f_{Z^2}(x) = \frac{1}{2\sqrt{x}} \left\{\phi(\sqrt{x}) + \phi(-\sqrt{x})\right\} \qquad \text{for } x \ge 0\,. \tag{28}$$

Plugging into (21) shows that

$$f_{Z^2}(x) = (2\pi x)^{-1/2} \exp(-x/2) \qquad \text{for } x \ge 0; \tag{29}$$

this is called the *Chisquare*(1) density. Note that Chisquare(1) is the same as Gamma(1/2, 1/2). Thus (20) shows that

$$\text{if } X_1, \ldots, X_n \text{ are i.i.d. } N(0,1), \text{ then } \sum_1^m X_i^2 \sim \text{Chisquare}(m) \tag{30}$$

where Chisquare$(m) \equiv$ Gamma$(m/2, 1/2)$.

Uniform and Related Distributions

If $f_X(x) = 1_{[a,b]}(x)/(b-a)$ for real numbers $-\infty < a < b < \infty$, then we say that $X \sim$ Uniform(a, b). By far the most important special case is Uniform$(0, 1)$. Note that if $U \sim$ Uniform$(0, 1)$, then $X \equiv (b-a)U + a \sim$ Uniform(a, b).

A generalization of this is the Beta(c, d) family. We say $X \sim$ Beta(c, d) if $f_X(x) = x^{c-1}(1-x)^{d-1}1_{[0,1]}(x)/B(c, d)$ where $B(c, d) = \Gamma(c)\Gamma(d)/\Gamma(c+d)$.

Suppose that $\xi_1, \ldots, \xi_n$ are i.i.d. Uniform$(0, 1)$. Let $0 \le \xi_{n:1} \le \ldots \le \xi_{n:n} \le 1$ denote the ordered values of the ξ_i's; we call the $\xi_{n:i}$'s the *uniform order statistics*. (Alternatively, if n is understood, then we also write $\xi_{(i)}$ for $\xi_{n:i}$, $i = 1, \ldots, n$.) It seems intuitive that $\xi_{n:i}$ equals x if $(i-1)$ of the ξ_i's fall in $[0, x)$, 1 of the ξ_i's is equal to x, and $n-i$ of the ξ_i's fall in $(x, 1)$. There are $n!/[(i-1)!(n-i)!]$ such designations of the ξ's, and the chance of the falling in the correct parts of $[0, 1]$ is $x^{i-1}(1-x)^{n-i}$. Thus

$$f_{\xi_{n:i}}(x) = \frac{n!}{(i-1)!(n-i)!} x^{i-1}(1-x)^{n-i} 1_{[0,1]}(x)\,; \tag{31}$$

in other words, $\xi_{n:i} \sim$ Beta$(i, n-i+1)$. Also note that the joint density of $(\xi_{n:1}, \ldots, \xi_{n:n})$ is given by

$$f_{\xi_{n:1},\ldots,\xi_{n:n}}(u_1, \ldots, u_n) = n! 1_A(u_1, \ldots, u_n) \tag{32}$$

where $A \equiv \{(u_1, \ldots, u_n) \in [0, 1]^n : \; 0 \le u_1 \le \ldots \le u_n \le 1\}$.

Exercise 1.10 Give a rigorous proof of (31) by computing $F_{\xi_{n:i}}$ and differentiating.

Exercise 1.11 Give a proof of (32).

The Cauchy Distribution

If $f_X(x) = \{b\pi[1+(x-a)^2/b^2]\}^{-1}$ on $(-\infty, \infty)$, then we say that $X \sim$ *Cauchy*(a,b). By far the most important special case is Cauchy$(0,1)$; in this case we say simply that $X \sim$ *Cauchy*, and its density is $[\pi(1+x^2)]^{-1}$ on $(-\infty,\infty)$. Verify that $E|X| = \infty$. We will see below that if $X_1, \ldots, X_n$ are i.i.d. Cauchy, then $\overline{X}_n \equiv (X_1 + \cdots + X_n)/n \sim$ Cauchy. These two facts make the Cauchy ideal for many counterexamples.

Rademacher Random Variables and Symmetrization

May problems become simpler if the problem is symmetrized. One way of accomplishing this is by the appropriate introduction of Rademacher random variables. We say that ϵ is a *Rademacher random variable* if $P(\epsilon = 1) = P(\epsilon = -1) = 1/2$. Thus $\epsilon \sim 2\,\text{Bernoulli}(1/2) - 1$.

We say that X is a *symmetric random variable* if $X \sim -X$. If X and X' are i.i.d., then $X^s \equiv (X - X') \sim (X' - X) = -(X - X') = -X^s$; hence X^s is a symmetric random variable.

Exercise 1.12 if X is a symmetric random variable independent of the Rademacher random variable ϵ, then $X \sim \epsilon X$.

The Multinomial Distribution

Suppose that $B_1 \cup \cdots \cup B_k = R$ for Borel sets B_i with $B_i \cap B_j = \emptyset$ for $i \neq j$; we call this a *partition* of R. Let $Y_1, \ldots, Y_n$ be i.i.d. random variables on $(\Omega, \mathcal{A}, P)$. Let $\underline{X}_i \equiv (X_{i1}, \ldots, X_{ik}) \equiv (1_{B_1}(Y_i), \ldots, 1_{B_k}(Y_i))$ for $i = 1, \ldots, n$, and set

$$(33)\qquad \underline{N} \equiv (N_1, \ldots, N_k) \equiv \sum_{i=1}^n \underline{X}_i = \left(\sum_{i=1}^n X_{i1}, \ldots, \sum_{i=1}^n X_{ik}\right) = \left(\sum_{i=1}^n 1_{B_1}(Y_i), \ldots, \sum_{i=1}^n 1_{B_k}(Y_i)\right).$$

Note that $X_{1j}, \ldots, X_{nj}$ are i.i.d. Bernoulli(p_j) with $p_j = P(Y_i \in B_j)$ and thus $N_j \sim$ Binomial(n, p_j) marginally. Note that $N_1, \ldots, N_k$ are *dependent* random variables; in particular, $N_1 + \cdots + N_k = n$. The joint distribution of $(N_1, \ldots, N_k)$ is called the *Multinomial*$(n, \underline{p})$ = *Multinomial*$_k(n, (p_1, \ldots, p_k))$ distribution. The number of ways to designate n_1 of the Y_i's to fall in B_1, …, n_k of the Y_i's to fall in B_k is the *multinomial coefficient*

$$(34)\qquad \binom{n}{n_1 \cdots n_k} \equiv \frac{n!}{n_1! \cdots n_k!} \qquad \text{where } n_1 + \cdots + n_k = n.$$

Each such designation occurs with probability $\prod_{i=1}^k p_i^{n_i}$. Hence

$$(35)\qquad P(\underline{N} = \underline{n}) = P(N_1 = n_1, \ldots, N_k = n_k) = \binom{n}{n_1 \cdots n_k} p_1^{n_1} \cdots p_k^{n_k}.$$

Now it is a trivial calculation that for $j \neq l$,

$$(36)\qquad Cov[X_{ij}, X_{il}] = E(1_{B_j}(Y_i)1_{B_l}(Y_i)) - E(1_{B_j}(Y_i))E(1_{B_l}(Y_i)) = -p_j p_l.$$

Thus

(37) $\quad Cov[N_j, N_l] = -np_jp_l \qquad \text{for} \quad j \neq l\,.$

Hence it follows that

$$(38) \qquad \begin{pmatrix} N_1 \\ \cdot \\ \cdot \\ \cdot \\ N_k \end{pmatrix} \sim \left(n \begin{pmatrix} p_1 \\ \cdot \\ \cdot \\ \cdot \\ p_k \end{pmatrix}, n \begin{pmatrix} p_1(1-p_1) & \cdot & \cdot & \cdot & -p_1p_k \\ \cdot & \cdot & & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot \\ -p_1p_k & \cdot & \cdot & \cdot & p_k(1-p_k) \end{pmatrix} \right).$$

Exercise 1.13 Consider, in the context of the multinomial distribution, two subsets $C = \cup_{i \in I} B_i$ and $D = \cup_{j \in J} B_j$ for some subsets $I, J \subset \{1, \ldots, k\}$, *not necessarily disjoint*, so that C and D are not necessarily disjoint either. Let

$$N_C = \sum_{i \in I} N_i, \quad p_C = \sum_{i \in I} p_i, \quad \hat{p}_C = N_C/n\,,$$

and

$$N_D = \sum_{j \in J} N_j, \quad p_D = \sum_{j \in J} p_j, \quad \hat{p}_D = N_D/n\,.$$

Compute $Cov[N_C, N_D]$. [Hint: it will involve the probability $p_{C \cap D} = P(Y_1 \in C \cap D)$.]

2 Convolution and Related Formulas

Convolution

If X and Y are independent random variables on $(\Omega, \mathcal{A}, P)$, then

$$
\begin{aligned}
F_{X+Y}(z) &= P(X+Y \le z) = \int\int_{x+y\le z} dF_X(x)dF_Y(y) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} dF_Y(y)dF_X(x) \\
&= \int_{-\infty}^{\infty} F_Y(z-x)dF_X(x) \equiv F_X \star F_Y(z) \qquad (1)
\end{aligned}
$$

is a formula called the *convolution formula*, for F_{X+Y} in terms of F_X and F_Y (the symbol $\star$ stands for convolution). In case X and Y have densities f_X and f_Y with respect to Lebesgue measure, then so does $X+Y$. In fact, since

$$
\begin{aligned}
\int_{-\infty}^{z}\int_{-\infty}^{\infty} f_Y(y-x)f_X(x)dxdy &= \int_{-\infty}^{\infty}\left\{\int_{-\infty}^{z} f_Y(y-x)dy\right\} f_X(x)dx \\
&= \int_{-\infty}^{\infty} F_Y(z-x)dF_X(x) = F_{X+Y}(z)\,,
\end{aligned}
$$

it follows from (1) that $X+Y$ has a density given by

(2) $$f_{X+Y}(z) = \int_{\infty}^{\infty} f_Y(z-x)f_X(x)dx \equiv f_Y \star f_X(z)\,.$$

Exercise 2.1 Use (2) to show that for X and Y independent:
(i) $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ implies $X+Y \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$.
(ii) $X \sim$ Cauchy$(0, \sigma_1)$ and $Y \sim$ Cauchy$(0, \sigma_2)$ implies $X+Y \sim$ Cauchy$(0, \sigma_1+\sigma_2)$.
(iii) $X \sim$ Gamma(r_1, θ) and $Y \sim$ Gamma(r_2, θ) implies $X+Y \sim$ Gamma(r_1+r_2, θ).

Exercise 2.2 (i) If $X_1, \ldots, X_n$ are i.i.d. $N(0,1)$, then $(X_1+\cdots+X_n)/\sqrt{n} \sim N(0,1)$.
(ii) If $X_1, \ldots, X_n$ are i.i.d. Cauchy$(0,1)$, then $(X_1+\cdots+X_n)/n \sim$ Cauchy$(0,1)$.

If X and Y are independent random variables taking values in $0, 1, 2, \ldots$, then clearly

(3) $$P(X+Y=k) = \sum_{i=0}^{k} P(X=i)P(Y=k-i) \qquad \text{for } \ k = 0, 1, 2, \ldots.$$

Exercise 2.3 Use (3) to show that for X and Y independent:
(i) $X \sim$ Poisson(λ_1) and $Y \sim$ Poisson(λ_2) implies $X+Y \sim$ Poisson$(\lambda_1+\lambda_2)$.
(ii) $X \sim$ Negative Binomial(m_1, p) and $Y \sim$ Negative Binomial(m_2, p) implies $X+Y \sim$ Negative Binomial(m_1+m_2, p).

A fundamental problem in probability theory is to determine constants a_n and $b_n > 0$ for which i.i.d. random variables $X_1, X_2, \ldots, X_n, \ldots$ satisfy

(4) $$(X_1+\cdots+X_n-a_n)/b_n \to_d G \qquad \text{as } \ n \to \infty$$

for some non-degenerate distribution G. Exercise 2 gives us two examples of such convergence; each was derived via the convlution formula. However, except in certain special cases, such as exercises

2.1 - 2.3, the various convolution formulas are too difficult to deal with directly, at least for $n-$fold convolutions for large n. For this reason we need a variety of central limit theorems. These will be stated in Chapter 2.

Other Formulas

Exercise 2.4 Suppose that X and Y are independent with $P(Y > 0) = 1$. Show that

$$(5) \qquad F_{XY}(z) \equiv P(XY \le z) = \int_0^\infty F_X(z/y)dF_Y(y) \qquad \text{for all} \ \ z\,,$$

$$(6) \qquad F_{X/Y}(z) \equiv P(X/Y \le z) = \int_0^\infty F_X(zy)dF_Y(y) \qquad \text{for all} \ \ z\,.$$

If F_X has a bounded density f_X and F_Y has a density f_Y (these are overly strong hypotheses), then F_{XY} and $F_{X/Y}$ have densities given by

$$(7) \qquad f_{XY}(z) = \int_0^\infty y^{-1} f_X(z/y) f_Y(y)dy \qquad \text{for all} \ \ z\,,$$

and

$$(8) \qquad f_{X/Y}(z) = \int_0^\infty y f_X(zy) f_Y(y)dy \qquad \text{for all} \ \ z\,.$$

Exercise 2.5 Let $X \sim N(0,1)$, $Y \sim \chi^2_m$, and $Z \sim \chi^2_n$ be independent. Show that

$$(9) \qquad \frac{X}{\sqrt{Y/m}} \sim \ \text{Student's} \ \ t_m \equiv t(m)\,,$$

$$(10) \qquad \frac{Y/m}{Z/n} \sim \ \text{Snedecor's} \ \ F_{m,n} = F(m,n)\,, \ \ \text{and}$$

$$(11) \qquad \frac{Y}{Y+Z} \sim \ \text{Beta}(m/2, n/2)$$

where

$$(12) \qquad f_{t(m)}(x) \equiv \frac{\Gamma((m+1)/2)}{\sqrt{\pi m}\,\Gamma(m/2)} \frac{1}{(1+x^2/m)^{(m+1)/2}} 1_{(-\infty,\infty)}(x)$$

and

$$(13) \qquad f_{F(m,n)}(x) = \frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)} \frac{(m/n)^{m/2} x^{m/2-1}}{(1+mx/n)^{(m+n)/2}} 1_{(0,\infty)}(x)\,.$$

Exercise 2.6 If $Y_1, \ldots, Y_{n+1}$ are i.i.d. Exponential(θ), then

$$(14) \qquad Z_i \equiv \frac{Y_1 + \cdots + Y_i}{Y_1 + \cdots + Y_{n+1}} \sim \ \text{Beta}(i, n-i+1)\,;$$

in other words the ratio on the left has the same distribution as the ith order statistic of a sample of n Uniform$(0,1)$ random variables.

Exercise 2.7 If $Y_1, \ldots, Y_{n+1}$ are i.i.d. Exponential(θ), as in Exercise 2.6, then the joint distribution of $(Z_1, \ldots, Z_n)$ is the same as that of the order statistics $(\xi_{n:1}, \ldots, \xi_{n:n})$ of n Uniform$(0,1)$ random variables.

3 The Multivariate Normal Distribution

We say that $Y = (Y_1, \ldots, Y_n)'$ is *jointly normal with* 0 *means* if there exist i.i.d. $N(0,1)$ random variables $X_1, \ldots, X_k$ and an $n \times k$ matrix A of known constants for which $Y = AX$. Note that the $n \times n$ *covariance matrix* Σ of Y is

(1) $\quad \Sigma = E(YY') = E(AXX'A') = AA'.$

Theorem 3.1 The following are equivalent:

(2) $\quad \Sigma$ is the covariance matrix of some random vector Y.

(3) $\quad \Sigma$ is symmetric and non-negative definite.

(4) $\quad$ There exists an $n \times n$ matrix A such that $\Sigma = AA'$.

Proof. (2) implies (3): Σ is symmetric since $E(Y_iY_j) = E(Y_jY_i)$. Also $a'\Sigma a = Var(a'Y) \geq 0$, so that $\Sigma \geq 0$.

(3) implies (4): Since Σ is symmetric, there exists an orthogonal matrix Γ such $\Gamma'\Sigma\Gamma = D$ with D diagonal. We let $a \equiv \Gamma b$, and since $\Sigma \geq 0$ we have

$$0 \leq a'\Sigma a = b'\Gamma'\Sigma\Gamma b = b'Db = \sum_{i=1}^{n} d_{ii} b_i^2$$

for all b, implying that all $d_{ii} \geq 0$. Thus

$$\Sigma = \Gamma D \Gamma' = \Gamma D^{1/2} D^{1/2} \Gamma' = (\Gamma D^{1/2})(\Gamma D^{1/2})' \equiv AA'$$

where $D^{1/2}$ denotes the diagonal matrix with entries $\sqrt{d_{ii}}$ on the diagonal.

(4) implies (2): Let $X_1, \ldots, X_n$ be i.i.d. $N(0,1)$. Let $X \equiv (X_1, \ldots, X_n)'$ and $Y = AX$. Then Y has covariance matrix $\Sigma = AA'$. $\square$

Theorem 3.2 If $Y = A^{n \times k} X^{k \times 1}$ where $X \sim N(0, I)$, then

(5) $\quad \phi_Y(t) \equiv Ee^{it'Y} = \exp\left(-\frac{1}{2}t'\Sigma t\right) \qquad \text{with } \Sigma \equiv AA'$

and $rank(\Sigma) = rank(A)$. Conversely, if $\phi_Y(t) = \exp(-t'\Sigma t/2)$ with $\Sigma \geq 0$ of rank k, then

(6) $\quad Y = A^{n \times k} X^{k \times 1} \qquad \text{with } rank(A) = k \text{ and } X \sim N(0, I).$

(Thus only rank(A) independent X_i's are needed.)

Proof. We use the fact that the characteristic function of a standard normal random variable X_j is $Ee^{itX_j} = \exp(-t^2/2)$ in the proof. Now

$$\begin{aligned}
\phi_Y(t) &= E\exp(it'AX) = E\exp(i(A't)'X) \\
&= \exp\left(-\frac{1}{2}(A't)'(A't)\right) \\
&= \exp\left(-\frac{1}{2}t'AA't\right)
\end{aligned}$$

where we used

$$\phi_X(t) = E\exp(it'X) = \exp(-|t|^2/2)$$

to get the third equality.

Conversely, suppose that $\phi_Y(t) = \exp(-t'\Sigma t/2)$ with $\mathrm{rank}(\Sigma) = k$. Then there exists an orthogonal matrix Γ such that

(a) $$\Gamma'\Sigma\Gamma = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}$$

where D is diagonal and $k \times k$. Let $Z = \Gamma' Y$ so that

$$\Sigma_Z = \Gamma'\Sigma\Gamma = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}.$$

Then

(b) $$\phi_Z(t) = \phi_Y(\Gamma t) = \exp(-t'\Gamma'\Sigma\Gamma t/2) = \prod_{i=1}^{k} \exp(-t_i^2 d_{ii}/2) \prod_{i=k+1}^{n} 1$$

so that $Z_1, \ldots, Z_k$ are independent $N(0, d_{11}), \ldots, N(0, d_{kk})$ and $Z_{k+1} = \cdots = Z_n = 0$. Let $X_i \equiv Z_i/\sqrt{d_{ii}} \sim N(0,1)$ for $i = 1, \ldots, k$ with $X_{k+1} \equiv \cdots \equiv X_n = 0$. Then

$$Y = \Gamma Z = \Gamma \begin{pmatrix} \sqrt{D} & 0 \\ 0 & 0 \end{pmatrix} X^{n\times 1} = \Gamma \begin{pmatrix} \sqrt{D} \\ 0 \end{pmatrix} \tilde{X}^{k\times 1} = A^{n\times k}\tilde{X}^{k\times 1}$$

with A of rank k. $\square$

Theorem 3.3 (i) If $Y = (Y_1, \ldots, Y_k, Y_{k+1}, \ldots, Y_n) \sim N_n(0, \Sigma)$ with

(7) $$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

then

(8) $$(Y_1, \ldots, Y_k)' \sim N_k(0, \Sigma_{11}).$$

(ii) If $\Sigma_{12} = 0$, then $(Y_1, \ldots, Y_k)'$ and $(Y_{k+1}, \ldots, Y_n)'$ are independent.
(iii) If $(Y_1, Y_2)'$ is jointly normal, then Y_1 and Y_2 are independent if and only if $Cov[Y_1, Y_2] = 0$.
(iv) Linear combinations of normals are normal.

Proof. (i) Use the first k coordinates of the representation $Y = AX$.
(ii) Use the fact that

(a) $$\phi_Y(t) = \exp\left(-\frac{1}{2}t'\begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix} t\right) = \exp(-t_1'\Sigma_{11}t_1/2)\exp(-t_2'\Sigma_{22}t_2/2),$$

which is the product of the characteristic functions of the marginal distributions.
(iii) Just apply (ii).
(iv) Now $Z^{m\times 1} \equiv B^{m\times n}Y^{n\times 1} = B(AX) = (BA)X$. $\square$

The preceding development can be briefly summarized by introducing the notation $X \sim N_n(0, I) \equiv N(0, I)$ and $Y \sim N_n(0, \Sigma)$. We will write $Y \sim N(\mu, \Sigma)$ if $Y - \mu \sim N(0, \Sigma)$. Note that P_Y is completely specified by μ and Σ. We call Y *non-degenerate*, and $Y_1, \ldots, Y_n$ will be called *linearly independent* if $\text{rank}(\Sigma) = n$. Of course

(9) $\quad Y$ is non-degenerate if and only if $\quad rank(A) = n\,.$

Exercise 3.1 Show that (Y_1, Y_2) can have normal marginals without being jointly normal. [Hint: consider starting with a joint $N(0, I)$ density on R^2 and move mass in a symmetric fashion to make the joint distribution non-normal, but still keeping the marginals normal.]

Theorem 3.4 If $Y \sim N(0, \Sigma)$ is nondegenerate, then Y has density (with respect to Lebesgue measure on R^n) given by

(10) $$f_Y(y) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2} y'\Sigma^{-1} y\right) \qquad \text{for all } \; y \in R^n\,.$$

Proof. Now $Y = AX$ where $AA' = \Sigma$, $\text{rank}(A) = n$, $|A| \neq 0$, $X \sim N(0, I)$. Hence for any Borel set $B_n \in \mathcal{B}_n$

(a) $$P(X \in B_n) = \int 1_{B_n}(x) f_X(x) dx = \int 1_{B_n}(x)\phi(x_1)\cdots\phi(x_n) dx_1 \cdots dx_n$$

where $f_X(x) = (2\pi)^{-n/2}\exp(-x'x/2)$. Since $X = A^{-1}Y$, for any Borel set B_n,

$$\begin{aligned}
P(Y \in B_n) &= P(AX \in B_n) = P(X \in A^{-1}B_n) = \int 1_{A^{-1}B_n}(x) f_X(x) dx \\
&= \int 1_{A^{-1}B_n}(A^{-1}y) f_X(A^{-1}y) \left|\frac{\partial x}{\partial y}\right| dy \\
&= \int 1_{B_n}(y)(2\pi)^{-n/2}\exp\left(-\frac{1}{2}(A^{-1}y)'(A^{-1}y)\right)\left|\frac{\partial x}{\partial y}\right| dy \\
&= \int_{B_n} (2\pi)^{-n/2}|\Sigma|^{-1/2}|\exp\left(-\frac{1}{2}y'\Sigma^{-1}y\right) dy
\end{aligned}$$

since $(A^{-1})'(A^{-1}) = (AA')^{-1} = \Sigma^{-1}$ and

(b) $$\left|\frac{\partial x}{\partial y}\right| = |A^{-1}| = \sqrt{|(A^{-1})'||A^{-1}|} = \sqrt{|\Sigma^{-1}|} = 1/\sqrt{|\Sigma|}\,.$$

□

Our last theorem about the multivariate normal distribution concerns the conditional distribution of one block of a joint normal random vector given a second block.

Theorem 3.5 If

(11) $$Y = \begin{pmatrix} Y^{(1)} \\ Y^{(2)} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right),$$

where $Y^{(1)}$ is a $k-$ vector, $Y^{(2)}$ is an $n-k$ - vector, and where Σ_{22} is nonsingular, then

(12) $$(Y^{(1)} \big| Y^{(2)}) \sim N_k((\mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(Y^{(2)} - \mu^{(2)})\,,\, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})\,.$$

Moreover, with $\Sigma_{11\cdot 2} \equiv \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$,

$$(13) \qquad \begin{pmatrix} Y^{(1)} - \mu^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}(Y^{(2)} - \mu^{(2)}) \\ Y^{(2)} - \mu^{(2)} \end{pmatrix} \sim N_n\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{11\cdot 2} & 0 \\ 0 & \Sigma_{22} \end{pmatrix} \right).$$

Proof. Without loss of generality, suppose that $\mu^{(i)} = 0$, $i = 1, 2$; otherwise subtract $\mu^{(i)}$ from $Y^{(i)}$, $i = 1, 2$. Then

$$(a) \qquad Z \equiv \begin{pmatrix} Z^{(1)} \\ Z^{(2)} \end{pmatrix} \equiv \begin{pmatrix} Y^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}Y^{(2)} \\ Y^{(2)} \end{pmatrix}$$

is just a linear combination of the Y_i's; and so it is normal, and all we need to know is μ_Z and Σ_Z. But $\Sigma_{Z,12} = \Sigma_{12} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22} = 0$, so that $Z^{(1)}$ and $Z^{(2)}$ are independent by Theorem 3.3. Also, $\Sigma_{Z,22} = \Sigma_{22}$ and

$$(b) \qquad \Sigma_{Z,11} = \Sigma_{11} - 2\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} + \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22}\Sigma_{22}^{-1}\Sigma_{21} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \Sigma_{11\cdot 2}\,.$$

Note that

$$(c) \qquad |\Sigma| = |\Sigma_{22}||\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}|\,.$$

Some Facts about Chi-Square Distributions

If $X \sim N_n(0, I)$, then $\|X\|^2 = X'X = \sum_{i=1}^n X_i^2 \sim \chi_n^2$, the Chisquare distribution with n degrees of freedom.

Corollary 1 If $Y \sim N_n(0, \Sigma)$ with Σ positive definite, then $Y'\Sigma^{-1}Y \sim \chi_n^2$.

Proof. $Y = AX$ where A is nonsingular and $X \sim N_n(0, I)$ and $\Sigma = AA'$. Hence $\Sigma^{-1} = (A')^{-1}A^{-1}$ and it follows that $Y'\Sigma^{-1}Y = X'A'(A')^{-1}A^{-1}AX = X'X \sim \chi_n^2$. □

Now for the noncentral Chisquare distributions: we will develop these in a series of steps as follows:

(a) Suppose that $X \sim N(\mu, 1)$. Define $Y \equiv X^2$, $\delta = \mu^2$. Then Y has density

$$(4) \qquad f_Y(y) = \sum_{k=0}^{\infty} p_k(\delta/2) g(y; (2k+1)/2, 1/2)$$

where $p_k(\delta/2) = \exp(-\delta/2)(\delta/2)^k/k!$, and $g(\cdot; (2k+1)/2, 1/2)$ is the Gamma$(2k+1)/2, 1/2)$ = Chisquare$(2k+1)$ density. Another way to say this is: $(Y|K = k) \sim \chi_{2k+1}^2$ where $K \sim$ Poisson$(\delta/2)$. We will say that Y has the *noncentral chisquare distribution with* 1 *degree of freedom and noncentrality parameter* δ, and write $Y \sim \chi_1^2(\delta)$ in this case.

(b) Now suppose that $X_1 \sim N(\mu, 1)$, and $X_2, \ldots, X_n \sim N(0, 1)$, and all of $X_1, \ldots, X_n$ are independent. Define $Y \equiv X'X = |X|^2$, $\delta = \mu^2$. Then Y has density

$$(5) \qquad f_Y(y) = \sum_{k=0}^{\infty} p_k(\delta/2) g(y; (2k+n)/2, 1/2)$$

where $p_k(\delta/2) = \exp(-\delta/2)(\delta/2)^k/k!$, and $g(\cdot;(2k+n)/2, 1/2)$ is the Gamma$((2k+n)/2, 1/2) =$ Chisquare$(2k+n)$ density. Another way to say this is: $(Y|K=k) \sim \chi^2_{2k+n}$ where $K \sim$ Poisson$(\delta/2)$. We will say that Y has the *noncentral chisquare distribution with n degrees of freedom and noncentrality parameter δ*, and write $Y \sim \chi^2_n(\delta)$ in this case.

(c) Now suppose that $X \sim N_n(\mu, I)$ and let $Y \equiv X'X$. Then $Y \sim \chi^2_n(\delta)$ with $\delta = \mu'\mu = |\mu|^2$.

Proof. Let Γ be an $n \times n$ orthogonal matrix with first row $\mu/|\mu| = \mu/\sqrt{\mu'\mu}$. Then $Z \equiv \Gamma X \sim N_n(\Gamma\mu, \Gamma\Gamma') = N_n((|\mu|, 0, \ldots, 0)', I)$, and hence by (b) above

(a) $\quad Y \equiv X'X = Z'\Gamma\Gamma'Z = Z'Z \sim \chi^2_n(\delta)$

with $\delta = |\mu|^2 = \mu'\mu$. $\square$

(d) Now suppose that $X \sim N_n(\mu, \Sigma)$ where Σ is nonsingular. Let $Y \equiv X'\Sigma^{-1}X$. Then $Y \sim \chi^2_n(\delta)$ with $\delta = \mu'\Sigma^{-1}\mu$.

Proof. Define $Z \equiv \Sigma^{-1/2}X$ where $\Sigma^{1/2}(\Sigma^{1/2})' = \Sigma$. Then $Z \sim N_n(\Sigma^{-1/2}\mu, I)$, so by (c),

(a) $\quad Y = X'\Sigma^{-1}X = Z'Z \sim \chi^2_n(\delta)$

with $\delta = \mu'(\Sigma^{-1/2})'\Sigma^{-1/2}\mu = \mu'\Sigma^{-1}\mu$. $\square$

Exercise 3.2 Verify (4).

Exercise 3.3 Verify (5).

4 Integration by Parts

Integration by Fubini's theorem or "integration by parts" formulas are useful in many contexts. Here we record a few of the most useful ones.

Proposition 4.1 Suppose that the left-continuous function U and the right-continuous function V are nondecreasing functions ($\uparrow$). Then for any $a \le b$

$$(1) \qquad U_+(b)V(b) - U(a)V_-(a) = \int_{[a,b]} U\,dV + \int_{[a,b]} V\,dU$$

and

$$(2) \qquad U(b)V(b) - U(a)V(a) = \int_{(a,b]} U\,dV + \int_{[a,b)} V\,dU$$

where $U_+(x) \equiv \lim_{y\downarrow x} U(y)$ and $V_-(x) \equiv \lim_{y\uparrow x} V(y)$.

Proof. We can apply Fubini's theorem 4.1.2 at steps (a) and (b) to obtain

$$\begin{aligned}
[U_+(b) - U(a)][V(b) - V_-(a)] &= \int_{[a,b]} \left\{ \int_{[a,b]} dU \right\} dV \\
\text{(a)} \qquad &= \int_{[a,b]} \int_{[a,b]} \{1_{x<y]}(x,y) + 1_{[x\ge y]}\} dU(x)dV(y) \\
\text{(b)} \qquad &= \int_{[a,b]} [U(y) - U(a)]dV(y) + \int_{[a,b]} [V(x) - V_-(a)]dU(x) \\
&= \int_{[a,b]} U\,dV - U(a)[V(b) - V_-(a)] + \int_{[a,b]} V\,dU - V_-(a)[U_+(b) - U(a)]\,.
\end{aligned}$$

A bit of algebra now gives (1). The proof of (2) is similar. $\square$

Mean, Variances, and Covariances

If $\xi \sim \text{Uniform}(0,1)$ and F is an arbitrary distribution function, then we will see in section 2.3 that $X \equiv F^{-1}(\xi)$ has distribution function F. Note note that this X satisfies

$$(3) \qquad X = \int_{(0,1)} F^{-1}(t)\,d1_{[\xi\le t]}$$

and

$$(4) \qquad X = \int_{(-\infty,\infty)} x\,d1_{[X\le x]}$$

where $1_{[\xi\le t]}$ is a random distribution function that puts mass 1 at the point $\xi(\omega)$ and $1_{[X\le x]}$ is a random distribution function that puts mass 1 at the point $X(\omega)$. If X has mean μ, then

$$(5) \qquad \mu = \int_{(0,1)} F^{-1}(t)\,dt = \int_{(-\infty,\infty)} x\,dF(x)\,.$$

Moreover, when μ is finite we can write

$$X - \mu = \int_{(0,1)} F^{-1}(t) d(1_{[\xi \le t]} - t) = -\int_{(0,1)} (1_{[\xi \le t]} - t) dF^{-1}(t) \tag{6}$$

or

$$X - \mu = \int_{(-\infty,\infty)} x d(1_{[X \le x]} - F(x)) = -\int_{(-\infty,\infty)} (1_{[X \le x]} - F(x)) dx \,. \tag{7}$$

The first formula in each of (6) and (7) is trivial; the second follows from integration by parts. For example, (6) is justified by

$$|tF^{-1}(t)| \le |\int_0^t F^{-1}(s) ds| \to 0 \qquad \text{as} \;\; t \to 0$$

when $E|X| = \int_0^1 |F^{-1}(t)| dt < \infty$, and the analogous result $(1-t)F^{-1}(t) \to 0$ as $t \to 1$. Thus when $Var(X) < \infty$, Fubini's theorem gives

$$\begin{aligned} Var(X) &= E\left\{ \int_{(0,1)} (1_{[\xi \le s]} - s) dF^{-1}(s) \int_{(0,1)} (1_{[\xi \le t]} - t) dF^{-1}(t) \right\} \\ &= \int_{(0,1)} \int_{(0,1)} E\{(1_{[\xi \le s]} - s)(1_{[\xi \le t]} - t)\} dF^{-1}(s) dF^{-1}(t) \\ &= \int_{(0,1)} \int_{(0,1)} (s \wedge t - st) dF^{-1}(s) dF^{-1}(t) \qquad (8) \end{aligned}$$

via (6), and the parallel formula

$$Var(X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [F(x \wedge y) - F(x)F(y)] dx dy \tag{9}$$

via (7). Of course we already know that

$$Var(X) = \int_0^1 [F^{-1}(t) - \mu]^2 dt = \int_{-\infty}^{\infty} (x - \mu)^2 dF(x) \,. \tag{10}$$

Now suppose that X, Y are random variables and let G, H denote measurable functions.

Proposition 4.2 (Formulas for means, moments, and covariances).
(i) If $X \ge 0$ has distribution function F, then

$$E(X) = \int_0^{\infty} (1 - F(x)) dx = \int_0^1 F^{-1}(t) dt \,. \tag{11}$$

(ii) If $E|X| < \infty$, then

$$E(X) = -\int_{-\infty}^0 F(x) dx + \int_0^{\infty} (1 - F(x)) dx = \int_0^1 F^{-1}(t) dt \,. \tag{12}$$

(iii) If $X \ge 0$, then

$$E(X^r) = r \int_0^{\infty} x^{r-1} (1 - F(x)) dx = \int_0^1 [F^{-1}(t)]^r dt \,. \tag{13}$$

(iv) If (X, Y) has joint distribution function F with marginal distribution functions F_X, F_Y, and G, H are nondecreasing, then

$$(14) \qquad Cov[G(X), H(Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [F(x,y) - F_X(x)F_Y(y)]dG(x)dH(y)\,.$$

Note the special case $G = H = I$ with $I(x) \equiv x$ for all $x \in R$.
(v) If K is $\uparrow$ and left continuous and $\xi \sim$ Uniform$(0,1)$ (perhaps $K = h(F^{-1})$ for an $\uparrow$ left continuous function h and for $X \equiv F^{-1}(\xi)$ for a distribution function F)

$$\begin{aligned}
(15) \qquad Var[K(\xi)] &= \int_0^1 \int_0^1 (s \wedge t - st)dK(s)dK(t) \\
(16) \qquad &= \int_{-\infty}^{\infty} \int_{\infty}^{\infty} [F(x \wedge y) - F(x)F(y)]dh(x)dh(y) \\
(17) \qquad &= Var[h(X)]\,.
\end{aligned}$$

(vi) If $X \geq 0$ is integer - valued,

$$(18) \qquad E(X) = \sum_{k=1}^{\infty} P(X \geq k)$$

and

$$(19) \qquad E(X^2) = \sum_{k=0}^{\infty} (2k+1)P(X > k)\,.$$

Exercise 4.1 Prove the formulas (11) - (13) using Fubini's theorem.

Exercise 4.2 Give an extension of (13) to arbitrary random variables in the case $r =$ an integer k.

Exercise 4.3 Prove formulas (14) and (15).

Exercise 4.4 For any distribution function F we have

$$\int [F(x+\theta) - F(x)]dx = \theta \qquad \text{for each} \quad \theta \geq 0\,.$$

Exercise 4.5 How should the left side of (1) be altered if we replace $[a, b]$ in both places on the right side of (1) by (a, b), or by $(a, b]$, or by $[a, b)$?