stats6251prob 625.1 - Probability background

Anonymous

December 1, 2015

Contents

1 Probability spaces and random variables

1.1 Probability background

1.1.1 Handout

Definition 1 A *probability space* consists of a set, Ω, the sample space (or population) with a collection $\mathcal{A}$ of sets called events A which are subsets of Ω (i.e. $A \subseteq \Omega$ so $\mathcal{A} \subseteq P(\Omega)$) and a probability measure which is a function

$$P : \mathcal{A} \to [0, 1]$$

satisfying the conditions $0 \leq P[A] \leq P[\Omega] = 1$ and

$$P[\bigcup_{i=1}^{\infty} A_i] = \sum_{i=1}^{\infty} P[A_i]$$

for $A_i \in \mathcal{A}$ such that $A_i \cap A_j = \emptyset$ if $i \neq j$

The Borel-algebra is the smallest collection of sets which contains the half-closed intervals, $[a, b[$, for $a, b \in \mathbb{R}, a < b$ (or appropriate subset of $\mathbb{R}$) and is closed with respect to countable unions and complements.

Along with the definition of random variables below, these formalities suffice for this course in mathematical statistics. Much more detail can be obtained in a course on measure theory or theoretical probability.

If A and B are events then the *probability of A given B* is

$$P[A|B] := \frac{P[A \cap B]}{P[B]}.$$

Suppose we have a bag of marbles. We observe two properties concerning the marbles. They are either green or yellow and either light or heavy. If we pull a marble out of the bag while blindfolded we can feel wether it is light or heavy. In this instance the marble was light. Knowing that we can now use the definition above to find the odds of the marble being green.

The event of the marble being light we call B, and the event of it being green A. Then we wish to find:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Because since we already know the weight of the marble the chances of the marble being green depend on the amount of green marbles of all marbles that are light. So we find the likelihood of a marble being green and light and divide by the likelihood of a marble being light.

1.2 Random variables

1.2.1 Handout

Definition 2 A *random variable* is a function

$$X : \Omega \to \mathbb{R}$$

such that $X^{-1}(B) \in \mathcal{A}$ if $B \in \mathcal{B}$, where $\mathcal{B}$ is the Borel-algebra over $\mathbb{R}$ so we can define

$$P[X \in B] = P[X^{-1}(B)].$$

Definition 3 The *cumulative distribution function* (cdf) is the function F defined by

$$F(x) := P[X \leq x].$$

Commonly an original sample space is not obvious but the possible outcomes of an experiment are in $\mathbb{R}$ and we define

$$X = id_{\mathbb{R}}$$

to obtain a random variable which has the desired probability distribution on $\mathbb{R}$.

Definition 4 A random variable X is *discrete* if $P[X = x] > 0$ for a countable number of x-values. In this case the *probability mass function* of X is the function

$$p(x) := P[X = x].$$

Definition 5 X is a *continuous* random variable if there is a function $f : \mathbb{R} \to \mathbb{R}_+$ such that

$$P[X \in A] = \int_A f(x)dx$$

for all events A.

Example 1 Consider two tosses of an unbiased coin. In this case the sample space is

$$\Omega = \{kk, ks, sk, ss\}$$

Where k means the result of the toss was heads, s means the result was tails.

$$X(\omega) = \begin{cases} 0 & \omega = kk, \\ 1 & \omega = ks \text{ or } sk, \\ 2 & \omega = ss. \end{cases}$$

If the coin being used is fair then $P(\omega) = 1/4$ for each $\omega \in \Omega$. Thus we can compute the chances of getting a certain amount of heads from our two tosses. If x is the number of heads then

x	$P[X = x]$
0	1/4
1	1/2
2	1/4

Example 2 The double-or-nothing game:

$$X_n := 2^n \chi_{[0,2^{-n}]}$$

The reader should elaborate and show that this represents a fair double-or-nothing game:

- What is Ω?
- What is P?
- Is it true that $P[X_{n+1} = 2X_n | X_n > 0] = 1/2$? Rewrite this in several ways.

Example 3

$$X_1, X_2, \ldots : [0,1] \to \{0,1\}$$

Split $[0,1[$ into the intervals

$$[\frac{k}{2^i}, \frac{k+1}{2^i}[$$

where $k = 0, 1, \ldots 2^i - 1$ and let

$$X_i(\omega) := \begin{cases} 0 & \frac{2j}{2^i} \leq \omega < \frac{2j+1}{2^i} \\ 1, \text{ otherwise.} \end{cases}$$

Then X_i, X_j are independent pairs if $i \neq j$.

Definition 6 Let X and Y be two random variables. The *Conditional mass function* of X given a value of the random variable Y is given by

$$P_{X|Y}(x|y) = P[X = x|Y = y] = \frac{P[X = x, Y = y]}{P[Y = y]} = \frac{P_{XY}(x,y)}{P_Y(y)}.$$

The *conditional density* of X given a value of the random variable Y is

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}, \quad f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}.$$

Example 4 Given that $P_{XY}(1,1) = 0.5$, $P_{XY}(2,1) = 0.1$, $P_{XY}(2,2) = 0.3$, $P_{XY}(1,2) = 0.1$, $P_Y(1) = 0.6$ calculate the probability of X=1 given that Y=1. We use the definition of the conditional mass function:

$$P_{X|Y}(1,1) = \frac{P_{XY}(1,1)}{P_Y(1)} = \frac{0.5}{0.6} = 5/6$$

1.3 Expected values

1.3.1 Handout

Definition 7 The *expected value* of a random variable X is

$$\mathbb{E}[X] := \begin{cases} \int x f(x) dx & \text{if this exists or more specifically if } \mathbb{E}[|X|] < \infty \\ \sum x p(x) \end{cases}$$

when f (p) if the density function (mass function) of X.

The *variance* of a random variable X is

$$Var[X] := \mathbb{E}\big[(X-\mu)^2\big]$$

when $\mu = \mathbb{E}[X]$ and all the integrals exist (and are finite).

2 Generating functions

2.1 Characteristic and moment generating functions

2.1.1 Handout

Definition 8 The *moment generating function* (m.g.f.) of the random variable X is the function

$$M_X(T) := \mathbb{E}\big[e^{tX}\big]$$

defined for those values of t where the expected value exists.

Definition 9 The *characteristic function* of (the distribution of) X is the function

$$\phi_X(t) := \mathbb{E}\big[e^{itX}\big]$$

Remark 2.1. ϕ_X always exists since

$$\mathbb{E}\big[|e^{itX}|\big] = \mathbb{E}[1] = 1$$

and hence both the real and imaginary parts of the integral exis so that $\mathbb{E}\big[e^{itX}\big]$ exists for $t \in \mathbb{R}$.

We will use the following result:

If $X_1, X_2, \ldots$ is a sequence of random variables with cumulative distribution functions F_n and characteristic functions ϕ_n such that $\phi_n(t) \to \phi(t)$ when $|t| < \varepsilon$ and ϕ corresponds to the cumulative distribution function F which is continuous at x, then $F_n(x) \to F(x)$. In other words,

$$P[X_n \le x] \to P[X \le x] \text{ if } \phi_n(t) \to \phi(t).$$

Example 5 If $X \sim G(\alpha, \beta)$ i.e. X has density

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, x > 0.$$

(the gamma density, discussed in detail later) then

$$\begin{aligned} M_X(t) &= \mathbb{E}\big[e^{tX}\big] = \int_0^\infty \frac{e^{tx} x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha} dx \\ &= \frac{\Gamma(\alpha)(\frac{-1}{t-1/\beta})^\alpha}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty \frac{x^{\alpha-1} e^{-\frac{x}{-1/(t-1/\beta)}}}{\Gamma(\alpha)(\frac{-1}{t-1/\beta})^\alpha} \\ &= \frac{1}{\beta^\alpha(\frac{1}{\beta}-t)^\alpha} = \frac{1}{(1-\beta t)^\alpha}. \end{aligned}$$

Theorem 2.1 Let $\varepsilon > 0$ and X be a random variable with moment generating function $M(t) = \mathbb{E}\big[e^{tX}\big]$ defined for $|t| < \varepsilon$. Then:

$$\mathbb{E}[X^n] = M^{(n)}(0) = \frac{\mathrm{d}^n}{dt^n} M(t)\bigg|_{t=0}.$$

Proof. If $M(t) = \int e^{tx} f(x)\,dx$ and if it is permissible to differentiate under the integral, then

$$M^{(n)}(t) = \int e^{tx} x^n f(x)\,dx \quad \text{and thus} \quad M^{(n)}(0) = \int x^n f(x)\,dx = \mathbb{E}[X^n].$$

Note also that if it is permissible to take the summation outside the expected value, then

$$\mathbb{E}\big[e^{tX}\big] = \mathbb{E}\Bigg[\sum_{n=0}^{\infty} \frac{(tX)^n}{n!}\Bigg] \stackrel{?}{=} \sum_{n=0}^{\infty} \mathbb{E}\bigg[\frac{t^n}{n!} X^n\bigg] = \sum_{n=0}^{\infty} \frac{t^n}{n!} \mathbb{E}[X^n],$$

so if $\mathbb{E}[X^n]$ exists and is limited for all n, then this is a "well-behaved" function and $M^{(n)}(0) = \mathbb{E}[X^n]$. ☐

Example 6 (a) **The standard normal distribution.** Let Z have the standard normal distribution, i.e. $Z \sim n(0,1)$ with density

$$f(\zeta) = \frac{1}{\sqrt{2\pi}} e^{-\zeta^2/2}, \quad \zeta \in \mathbb{R}.$$

The cumulative distribution function is

$$F(\zeta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\zeta} e^{-t^2/2}\,dt, \quad \zeta \in \mathbb{R},$$

and the moment generating function is

$$\begin{aligned} M_Z(t) &= \int e^{tx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx \\ &= \frac{1}{\sqrt{2\pi}} \int e^{-\frac{1}{2}(x^2 - 2tx)}\,dx \\ &= e^{\frac{1}{2}t^2} \cdot \frac{1}{\sqrt{2\pi}} \int e^{-\frac{1}{2}(x-t)^2}\,dx \\ &= e^{\frac{1}{2}t^2}, \quad t \in \mathbb{R}. \end{aligned}$$

We thus obtain

$$M_Z'(t) = t e^{\frac{1}{2}t^2} \quad \text{og} \quad M_Z''(t) = e^{\frac{1}{2}t^2} + t^2 e^{\frac{1}{2}t^2},$$

and from the previous theorem it follows that

$$\mathbb{E}[Z] = M_Z'(0) = 0 \quad \text{og} \quad \mathbb{E}\big[Z^2\big] = M_Z''(0) = 1.$$

Finally we have

$$\mathrm{Var[Z]} = \mathbb{E}\big[(\mathrm{Z} - \mu)^2\big] = \mathbb{E}\big[\mathrm{Z}^2 - 2\mathrm{Z}\mu + \mu^2\big] = \mathbb{E}\big[\mathrm{Z}^2\big] - (\mathbb{E}[\mathrm{Z}])^2 = 1.$$

(b) **The general normal distribution.** Let $X := \sigma Z + \mu$ with $Z \sim n(0,1)$. Then clearly $\mathbb{E}[X] = \sigma\mathbb{E}[Z] + \mu = \mu$ and

$$\begin{aligned}\mathrm{Var[X]} &= \mathbb{E}\big[X^2\big] - (\mathbb{E}[X]^2) \\ &= \mathbb{E}\big[(\sigma Z + \mu)^2\big] - \mu^2 \\ &= \mathbb{E}\big[\sigma^2 Z^2 + 2\sigma\mu Z + \mu^2\big] - \mu^2 \\ &= \sigma^2\mathbb{E}\big[Z^2\big] + 2\sigma\mu\mathbb{E}[Z] + \mu^2 - \mu^2 \\ &= \sigma^2.\end{aligned}$$

The r.v. X is said to have a **general normal distribution** with expected value μ and variance σ^2, denoted $X \sim n(\mu, \sigma^2)$. The moment generating function is

$$M_X(t) = \mathbb{E}\Big[e^{t(\sigma Z+\mu)}\Big] = \mathbb{E}\big[e^{t\sigma Z + t\mu}\big] = e^{t\mu}\mathbb{E}\Big[e^{(t\sigma)Z}\Big] = e^{t\mu}M_Z(t\sigma), \quad t \in \mathbb{R}.$$

The c.d.f of the random variable is given by

$$F_X(x) = \mathbb{P}(X \le x) = \mathbb{P}(\sigma Z + \mu \le x) = \mathbb{P}\big(Z \le \tfrac{x-\mu}{\sigma}\big) = F_Z(\tfrac{x-\mu}{\sigma}), \quad x \in \mathbb{R},$$

and its density is therefore

$$f_X(x) = \tfrac{\mathrm{d}}{dx}F_X(x) = \tfrac{\mathrm{d}}{dx}F_Z(\tfrac{x-\mu}{\sigma}) = \tfrac{1}{\sigma}f_Z(\tfrac{x-\mu}{\sigma}) = \tfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

Theorem 2.2 Let $\varepsilon > 0$ and $X_1, X_2, \dots$ be random variables with moment generating functions $M_{X_1}, M_{X_2}, \dots$ such that $M_{X_n}(t) \to M(t)$, $n \to \infty$, fyrir $|t| < \varepsilon$. If M is the moment generating function of the random variable X, then $F_{X_n}(x) \to F_X(x)$ for all x where F_X is continuous.

Theorem 2.3 Let $X_1, \dots, X_n$ be independent random variables with moment generating functions $M_{X_1}, \dots, M_{X_n}$ and, as before $\bar{X} := \frac{1}{n}\sum_{i=1}^n X_i$ to obtain:

$$M_{\bar{X}}(t) = \prod_{i=1}^n M_{X_i}(t/n) \quad \text{og} \quad M_{\sum X_i}(t) = \prod_{i=1}^n M_{X_i}(t).$$

In particular, if $X_1, \dots, X_n$ all have the same moment generating function M:

$$M_{\bar{X}}(t) = (M(t/n))^n \quad \text{og} \quad M_{\sum X_i}(t) = (M(t))^n.$$

Example 7 Let $X_1, \dots, X_n \sim \text{Gamma}(\alpha, \beta)$ be independent with $\alpha, \beta > 0$ so each X_i has the density

$$f_{X_i}(x) = \frac{x^{\alpha-1}e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha}, \quad x > 0,$$

and moment generating function

$$M(t) = \frac{1}{(1-\beta t)^\alpha}.$$

From the above theorem we see that

$$M_{\bar{X}}(t) = (M(t/n))^n = \left(1 - \beta \frac{t}{n}\right)^{-n\alpha} = \frac{1}{\left(1 - \frac{\beta}{n} t\right)^{n\alpha}},$$

which implies that $X \sim \text{Gamma}(\text{n}\alpha, \beta/\text{n})$. In addition

$$M_{\sum X_i}(t) = (M(t))^n = \left(\frac{1}{(1-\beta t)^{\alpha}}\right)^n = \frac{1}{(1-\beta t)^{n\alpha}},$$

which shows that $\sum_{i=1}^n X_i \sim \text{Gamma}(\text{n}\alpha, \beta)$.

3 On multivariate transforms

3.1 Background to some multivariate transformations

3.1.1 Handout

Before going further we need some results from calculus of several variables. First recall that if the function

$$\mathbf{g} : \mathbb{R}^m \to \mathbb{R}^n; \quad \mathbf{g} := (g_1, \ldots, g_n)'$$

is one-to-one and continuously differentiable then the Jacobian determinant of the transformation is given by

$$J = \left|\frac{\partial \mathbf{g}}{\partial \mathbf{x}}\right| = |\nabla g_1 \ \cdots \ \nabla g_n| = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_n}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_1}{\partial x_m} & \cdots & \frac{\partial g_n}{\partial x_m} \end{vmatrix}.$$

For "convenient" regions $R \subseteq \mathbb{R}^n$ and a function $\mathbf{f}$ which is continuous on $\mathbf{g}(R)$ we have

$$\int_{\mathbf{g}(R)} \mathbf{f}(\mathbf{x}) \mathrm{d}\mathbf{x} = \int_R \mathbf{f}(\mathbf{g}(\mathbf{u})) |J| \mathrm{d}\mathbf{u}.$$

We therefore see that if $\mathbf{U}$ is a random variable with $\mathbf{X} = \mathbf{g}(\mathbf{U})$, then

$$f_{\mathbf{U}}(\mathbf{u}) = f_{\mathbf{X}}(\mathbf{g}(\mathbf{u})) |J|.$$

Example 8 Let X and Y be continuous and independent random variables and define $Z := X + Y$. If $W := X$, and consider the transformation

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} w \\ \zeta \end{pmatrix} := \begin{pmatrix} x \\ x + y \end{pmatrix}$$

where $J = \left|\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right| = 1$, and from the above we see that

$$f_{W,Z}(w, \zeta) = f_{X,Y}(w, \zeta - u)|J| = f_{X,Y}(w, \zeta - u) = f_X(w) f_Y(\zeta - u).$$

Hence we see that the marginal density function of Z is given by

$$f_Z(\zeta) = \int_{-\infty}^{\infty} f_{W,Z}(w, \zeta)\, dw = \int_{-\infty}^{\infty} f_X(u) f_Y(\zeta - u)\, du.$$

This can be derived in several different ways, e.g.

$$\begin{aligned} F_Z(\zeta) &= \mathbb{P}(Z \le \zeta) \\ &= \mathbb{P}(X + Y \le \zeta) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\zeta - x} f(x, y)\, dx\, dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\zeta - x} f_X(x) f_Y(y)\, dx\, dy \\ &= \int_{-\infty}^{\infty} f_X(x) F_Y(\zeta - x)\, dx. \end{aligned}$$

Example 9 Let $X \sim \text{Cauchy}(0, 1)$ with density

$$f_X(x) = \frac{1}{\pi} \frac{1}{1 + x^2}, \quad x \in \mathbb{R}.$$

For this random variable we see that

$$\mathbb{E}[|X|] = \int_{-\infty}^{\infty} \frac{|x|}{\pi(1 + x^2)}\, dx = 2 \int_0^{\infty} \frac{x}{\pi(1 + x^2)}\, dx = \infty,$$

and hence the expected value $\mathbb{E}[X]$ is not defined.

We say that X has **a general Cauchy-distribution** with parameters μ and σ^2, denoted $X \sim \text{Cauchy}(\mu, \sigma^2)$, if it has the density

$$f_X(x) = \frac{1}{\pi\sigma} \frac{1}{1 + (\frac{x-\mu}{\sigma})^2}, \quad x \in \mathbb{R}.$$

Recall that if X_1 and X_2 are independent random variables and $\text{Var}[X_1] = \text{Var}[X_2] = \sigma^2$, then

$$\text{Var}\left[\frac{X_1 + X_2}{2}\right] = \frac{\text{Var}[X_1] + \text{Var}[X_2]}{4} = \frac{\sigma^2}{2}$$

and in general we have that if $X_1, \dots, X_n$ are independent random variables and $\text{Var}[X_i] = \sigma^2$, then

$$\text{Var}\left[\frac{X_1 + \cdots + X_n}{n}\right] = \frac{\sigma^2}{n}.$$

because:

$$\text{Var}\left[\frac{X_1 + \cdots + X_n}{n}\right] = \text{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1^2}{n^2}\text{Var}\left[\sum_{i=1}^{n} X_i\right] = \frac{1^2}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

Example 10 On the other hand if $X_1, X_2 \sim \text{Cauchy}(0, 1)$ are independent, then

$$\frac{X_1 + X_2}{2} \sim \text{Cauchy}(0, 1)$$

Let's derive the result:

Let $X_1, X_2 \sim$ Cauchy(0,1) iid. and define $Z := \frac{X_1+X_2}{2}$. The pdf of a $X \sim$ Cauchy(0,1) is $f_X(x) = \frac{1}{\pi}\frac{1}{1+x^2}$.

It is known that $E[X] = \infty$ so the mgf for the Cauchy distribution doesn't exist. However the characteristic function does exist, defined by $\phi_X(t) = E[e^{itX}]$, $t \in \mathbb{R}$.

If we can show that $\phi_Z(t) = \phi_X(t)$ then it follows that the variables have the same distribution function, $F_Z(X) = F_X(X)$, and thus follow the same distribution i.e. $Z \sim$ Cauchy(0,1).

Let's begin with finding $\phi_X(t)$:

$$\phi_X(t) = E[e^{itX}] = \int_{-\infty}^{+\infty} e^{itX} f_X(x)dx = \int_{-\infty}^{+\infty} e^{itX}\frac{1}{\pi}\frac{dx}{1+x^2} \tag{1}$$

We use contour integration to calculate this integral. Define a closed path $\gamma :=$ $<$-R,$R$$>$ $* \beta_R$ where β_R is a half circle from R to $-R$ in the upper plane H_+. Let $g(z) = \frac{e^{itz}}{1+z^2}$ and integrate it along γ. So by the residue theory we get

$$\pi\phi_X(t) = \int_\gamma g(z)dz = \int_{<-R,R>} g(z)dz + \int_{\beta_R} g(z)dz = 2\pi i \sum_{\alpha_j \in H_+} \text{Res}(g, \alpha_j) \tag{2}$$

where α_j are poles of $g(z)$ in the upper half plane.

Let's show that $\int_{\beta_R} g(z)dz \to 0$ as $R \to \infty$:

$$\begin{aligned}\left|\int_{\beta_R} g(z)dz\right| &\le \int_{\beta_R} |g(z)||dz| \\ &= \int_{\beta_R} \frac{|e^{itz}|}{|1+z^2|} \\ &\le \int_{\beta_R} \frac{|dz|}{|1+z^2|} \\ &\le \sup_{|z|=R} \frac{1}{|1+z^2|}\int_{\beta_R} |dz| \\ &\le \frac{\pi R}{R^2-1} \to 0 \text{ as } R \to \infty\end{aligned}$$

Since $g(z)$ has poles of order 1 at $\alpha_1 = i \in H_+$ and $\alpha_2 = -i \in H_-$. The residue at α_1 is

$$\text{Res}(g, i) = \lim_{z \to i}(z-i)g(z) = \lim_{z \to i}(z-i)\frac{e^{itz}}{(z-i)(z+i)} = \frac{e^{-|t|}}{2i} \tag{3}$$

Note the $|t|$ since $t \in \mathbb{R}$.

Take the limit of (2) as $R \to \infty$ and get

$$\pi\phi_X(t) = 2\pi i \frac{e^{-|t|}}{2\pi} = \pi e^{-|t|}$$

and so

$$\phi_X(t) = e^{-|t|} \tag{4}$$

Let's find the characteristic function of Z:

$$\begin{aligned}\phi_Z(t) &= \phi_{\frac{X_1+X_2}{2}}(t) \\ &= E\left[e^{\frac{it(X_1+X_2)}{2}}\right] = E\left[e^{\frac{itX_1}{2}}e^{\frac{itX_2}{2}}\right] \\ &= E\left[e^{\frac{itX_1}{2}}\right]E\left[e^{\frac{itX_2}{2}}\right] = \phi_{X_1}\left(\frac{t}{2}\right)\phi_{X_2}\left(\frac{t}{2}\right) \\ &= e^{-\left|\frac{t}{2}\right|}e^{-\left|\frac{t}{2}\right|} = \left(e^{-\left|\frac{t}{2}\right|}\right)^2 = e^{-|t|}\end{aligned}$$

Thus we have shown that $\phi_{X_1}(t) = \phi_{X_2}(t) = \phi_Z(t)$ and thereby it follows that $F_{X_1} = F_{X_2} = F_Z$ and so $Z \sim$ Cauchy(0,1).

More generally if $X_1, \ldots, X_n \sim \text{Cauchy}(0,1)$ then

$$\frac{X_1 + \ldots + X_n}{n} \sim \text{Cauchy}(0,1).$$

Theorem 3.1 (Property of mean and variance of normals) Let $X_1, \ldots, X_n \sim n(\mu, \sigma^2)$ be independent random variables and define

$$\bar{X} := \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{og} \quad S^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

then:

(i) $\bar{X}$ and S^2 are independent random variables.

(ii) $\bar{X} \sim n(\mu, \sigma^2/n)$.

(iii) $\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}$.

Proof. to be done... □

4 The gamma, chi-square and t distributions

4.1 Gamma, chisquare and t

4.1.1 Handout

Example 11 Let $\alpha, \beta > 0$ and $x > 0$. Then $\frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^\alpha}$ is a probability density function:

$$\begin{aligned}\frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x^{\alpha-1}e^{-\frac{x}{\beta}}dx &= \frac{\beta^\alpha}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty y^{\alpha-1}e^{-y}dy \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)\beta^\alpha}\cdot\Gamma(\alpha) = 1\end{aligned}$$

where we substitute $y = \frac{x}{\beta}$ to get the first equality, and the second equality follows from the definition of the gamma function.

Definition 10 *The density of the gamma distribution is given by*

$$\frac{x^{\alpha-1}e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha},\ x > 0$$

and moment generating function

$$M(t) = (1-\beta t)^{-\alpha},\ t < \frac{1}{\beta}.$$

In the case of $\alpha = \nu/2,\ \beta = 2$ this is called a χ^2 - distribution with ν degrees of freedom and density

$$\frac{x^{\nu/2-1}e^{-x/2}}{\Gamma(\frac{\nu}{2})2^{\nu/2}},\ x > 0.$$

Example 12 *The mean of the gamma distribution is given by*

$$E(X) = \int_0^\infty xf(x)dx$$

$$\int_0^\infty x\frac{x^{\alpha-1}e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha}dx$$

$$\int_0^\infty \frac{x^{\alpha}e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha}dx$$

$$\frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x^\alpha e^{-x/\beta}dx$$

Substitute $x = u\beta, dx = \beta du$ to get

$$\frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty u^\alpha\beta^\alpha e^{-u}\beta du$$

$$\frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty u^\alpha\beta^{\alpha+1} e^{-u}du$$

$$\frac{\beta^{\alpha+1}}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty u^\alpha e^{-u}du$$

This then simplifies and due to the fact

$$\int_0^\infty u^\alpha e^{-u}du = \Gamma(\alpha+1)$$

We get

$$\frac{\beta\Gamma(\alpha+1)}{\Gamma(\alpha)}$$

Due to $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ We get $E(X) = \alpha\beta$ as the mean of the gamma distribution.

Example 13 *For $Z^2 \sim n\,(0,1)$ it is easy to that $Z^2 \sim \chi_1^2$*

Find the distribution of $X = Z^2$, where

$$f(z) = \frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-\mu)}{2\sigma^2}}$$

Lets begin with the cdf of X

$$F_X(x) = P(X \le x) = P(Z^2 \le x) = P(-\sqrt{x} \le Z \le \sqrt{x})$$

From this we get

$$F_X(x) = F_Z(-\sqrt{x}) - F_Z(\sqrt{x})$$

And finally we have:

$$f_X(x) = \frac{1}{2}x^{\frac{-1}{2}}\frac{1}{\sqrt{2\pi}}e^{\frac{-x}{2}} + \frac{1}{2}x^{\frac{-1}{2}}\frac{1}{\sqrt{2\pi}}e^{\frac{-x}{2}} = \frac{1}{2^{\frac{1}{2}}\sqrt{2\pi}}x^{\frac{-1}{2}}e^{\frac{-x}{2}}$$

This is the pdf of $\Gamma(\frac{1}{2}, 2)$ and is called the chi-square distribution with 1 degree of freedom, that is $Z^2 \sim \chi_1^2$

- Using the moment generating function we see that the sum of independent gamma random variables (with the same β) is a gamma-distributed random variable.
- We therefore also see that if $z_1, ..., z_n \sim n(0,1)$ iid then

$$z_1^2 + ... + z_n^2 \sim \chi_n^2.$$

Example 14 If $X \sim \chi_\nu^2$, then $\mathbb{E}[[X]] = \nu$. The probability density function of X is

$$f_X(x) = \begin{cases} cx^{\left(\frac{\nu}{2}-1\right)}e^{-\frac{1}{2}x}, & \text{if } x \geq 0. \\ 0, & \text{otherwise} \end{cases}$$

where $c = 2^{\frac{n}{2}}\Gamma\left(\frac{\nu}{2}\right)$ and $\Gamma()$ is the gamma function.

By definition: $E[X] = \int_0^\infty xf_X(x)dx$

From that we get:

$$E[X] = \int_0^\infty xcx^{\left(\frac{\nu}{2}-1\right)}e^{-\frac{1}{2}x}dx$$

$$E[X] = c\int_0^\infty x^{\left(\frac{\nu}{2}-1+1\right)}e^{-\frac{1}{2}x}dx$$

$$E[X] = c([-x^{\left(\frac{\nu}{2}\right)}2e^{-\frac{1}{2}x}]_{x=0}^\infty + \int_0^\infty \frac{\nu}{2}x^{\left(\frac{\nu}{2}-1\right)}2e^{-\frac{1}{2}x}dx)$$

$$E[X] = c((0-0) + \nu\int_0^\infty x^{\left(\frac{\nu}{2}-1\right)}e^{-\frac{1}{2}x}dx)$$

$$E[X] = \nu\int_0^\infty cx^{\left(\frac{\nu}{2}-1\right)}e^{-\frac{1}{2}x}dx$$

$$E[X] = \nu\int_0^\infty xf_X(x)dx$$

By definition: $\int_0^\infty f_X(x)dx = 1$ because $f_X(x)$ is a pdf. From that we get:

$$E[X] = \nu$$

Example 15 If $V \sim \chi_v$ then $Var[V] = 2v$

Let $X \sim \chi_n$ The probability density function of X is

$$f_X(x) = \begin{cases} cx^{\left(\frac{n}{2}-1\right)}e^{-\frac{1}{2}x}, & \text{if } x \geq 0. \\ 0, & \text{otherwise} \end{cases}$$

where $c = 2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)$ and $\Gamma()$ is the gamma function.

We know that $Var[X] = E\left[X^2\right] - (E[X])^2$. Now:

$$E\left[X^2\right] = \int_0^\infty x^2 f_X(x)dx$$

$$= \int_0^\infty x^2 cx^{n/2-1}e^{-x/2}dx$$

$$= c\int_0^\infty x^{n/2+1}e^{-x/2}dx$$

integration by parts:

$$= c\left[-x^{n/2+1}2e^{-x/2}\right]_{x=0}^\infty + \int_0^\infty \left(\frac{n}{2}+1\right)x^{n/2}2e^{-x/2}dx$$

$$= c(n+2)\int_0^\infty x^{n/2}e^{-x/2}dx$$

integration by parts:

$$= c(n+2)\left[-x^{n/2}2e^{-x/2}\right]_{x=0}^\infty + \int_0^\infty \frac{n}{2}x^{n/2-1}2e^{-x/2}dx$$

$$= c(n+2)\left(n\int_0^\infty x^{n/2-1}e^{-x/2}dx\right)$$

$$= (n+2)n\int_0^\infty cx^{n/2-1}e^{-x/2}dx$$

$$= (n+2)n\int_0^\infty f_X(x)dx$$

integral of the pdf over the support $[0,\infty)$ equals 1:

$$= (n+2)n$$

$$= n^2 + 2n$$

$E\left[X\right]^2 = n^2$

Now it's clear to see that $Var[X] = n^2 + 2n - n^2 = 2n$

Definition 11 If $Z \sim n(0,1)$ and $V \sim \chi^2_\nu$, then the distribution of the random variable $Z/\sqrt{V/\nu}$ is termed the *t-distribution with ν degrees of freedom*, denoted $T \sim t_\nu$.

We can fins the density of T by considering the function $(U,V) \mapsto (T,W)$ with $W := V$, thus obtaining the joint density of T and W and then integrating out W.

Definition 12 If $U \sim \chi^2_{\nu_1}$ and $V \sim \chi^2_{\nu_2}$ then the distribution of the random variable

$$\frac{U/\nu_1}{V/\nu_2}$$

is termed the *F-distribution with ν_1 and ν_2 degrees of freedom.* denoted $F \sim F_{\nu_1,\nu_2}$.

We have a general interest in drawing conclusions about μ when $X_1, ..., X_n \sim n(\mu, \sigma^2)$ are independent but μ, σ^2 are all unknown numbers. Such conclusions always build on the fact that

$$\bar{X} := \frac{1}{n}\sum_{i=1}^{n} X_i \sim n\left(\mu, \frac{\sigma^2}{n}\right)$$

so that

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim n(0, 1)$$

and if

$$S := \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

then

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

which are independent of $\bar{X}$, and hence

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{\frac{\bar{X}-\mu}{\sigma/n}}{\sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2/\sigma^2}{n-1}}} \sim t_{n-1}.$$

A consequence of this is that if $\mu = \mu_0$ then the number $t := \frac{\bar{x}-\mu}{s/\sqrt{n}}$ will in 95% of all experiments be between $2,5\%$ and $97,5\%$ probability limits in the t-distribution.

5 Linear combinations of random variables

5.1 General linear combinations

5.1.1 Handout

Recall that if X and Y are random variables with expected value

$$\mu_X = \mathbb{E}[X] \quad and \quad \mu_Y = \mathbb{E}[Y],$$

then the **covariance** of X and Y is defined by

$$Cov(X, Y) := \mathbb{E}[(X - \mu_X)(Y - \mu_Y)].$$

Special case: $X = Y \Rightarrow Cov(X, Y) = \mathrm{Var[X]} = \sigma^2_{\mathrm{X}}$ - if this expected value exists. Also recall that if X and Y are independent, then $Cov(X, Y) = 0$ since it is easy to see that

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) \Rightarrow Cov(X, Y) = \int\int (x - \mu_X)(y - \mu_Y)f_X(x)f_Y(y)dxdy = 0.$$

Theorem 5.1 If $X_1, ..., X_n$ are random variables and $Y_1, ..., Y_m$ are random variables with $Cov(X_i, Y_j) = \sigma_{ij}$ and $a_1, ..., a_n, b_1, ..., b_m$ are real numbers, then

$$\text{Cov}(\mathbf{a}'\mathbf{X}, \mathbf{b}'\mathbf{Y}) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \sigma_{ij}.$$

Proof. We now have

$$\begin{aligned}
\text{Cov}\left(\mathbf{a}'\mathbf{X}, \mathbf{b}'\mathbf{Y}\right) &= \mathbb{E}\Big[\Big(\sum a_i X_i - E\sum a_i X_i\Big)\Big(\sum b_j Y_j - E\sum b_j Y_j\Big)\Big] \\
&= \mathbb{E}\Big[\Big(\sum a_i X_i - \sum a_i E X_i\Big)\Big(\sum b_j Y_j - \sum b_j E Y_j\Big)\Big] \\
&= \mathbb{E}\left[\{\sum_{i=1}^{n} a_i\left(X_i - EX_i\right)\}\{\sum_{j=1}^{m} b_j\left(Y_j - EY_j\right)\}\right] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m}\mathbb{E}[a_i\left(X_i - EX_i\right) b_j\left(Y_j - EY_j\right)] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \sigma_{ij}.
\end{aligned}$$

as required. □

Definition 13 *The variance-covariance matrix* of the random variables (or random vector) $(X_1, ..., X_n)$ is the matrix

$$\mathbf{\Sigma} = (\sigma_{ij}) = (Cov(X_i, X_j)).$$

Corrollary 5.1 If $X_1, ...X_n$ are s.t. $Cov(X_i, X_j) = 0$ if $i \neq j$ and $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, then $\text{Cov}(\mathbf{a}'\mathbf{X}, \mathbf{b}'\mathbf{X}) = \sum_{i=1}^{n} a_i b_i \sigma_i^2$ $[= (\mathbf{a}'b)\sigma^2$ if $\sigma_i^2 = \sigma^2\ \forall i]$.

Corrollary 5.2 If $X_1, ..., X_n$ are such that $\sigma_{ij} = \delta_{ij}\sigma^2$ and $\mathbf{a}, \mathbf{b} \in \mathbb{R}$ are such that $\mathbf{a} \perp \mathbf{b}$, then $\text{Cov}(\mathbf{a}'\mathbf{X}, \mathbf{b}'\mathbf{X}) = 0$.

Corrollary 5.3 If $(X_1, ..., X_n)'$ is a vector r.v. with $\mathbb{E}[\mathbf{X}] = \mu, \text{Var}[\mathbf{X}] = \text{Cov}(\mathbf{X}) = \mathbf{\Sigma}$ and $\mathbf{a} \in \mathbb{R}^n$, then $E\mathbf{a}'\mathbf{X} = \mathbf{a}'\mu$ and $V\mathbf{a}'\mathbf{X} = \mathbf{a}'\mathbf{\Sigma}\mathbf{a}$.

Corollary 5.4 $\text{Cov}(\mathbf{a}'\mathbf{X}, \mathbf{b}'\mathbf{X}) = \mathbf{a}'\mathbf{\Sigma}\mathbf{b}$.

Corollary 5.5 X vector r.v., $E\mathbf{X} = \mu$, $V\mathbf{X} = \mathbf{\Sigma}$. A is an $n \times n$ matrix, then $\mathbb{E}[AX] = A\mu$ og $\text{Var[AX]} = \text{A}\mathbf{\Sigma}\text{A}^{\text{T}}$.

5.2 Linear combinations of Gaussian random variables

5.2.1 Handout

Theorem 5.2 Let $X_1, ..., X_n \sim n(0,1)$ be independent, let $X = (X_1, ..., X_n)'$ and let Y be the r.v. $\mathbf{Y} := P\mathbf{X} + \mu$ where P is a matrix with $\text{rank}(P) = n$ and $\mu \in \mathbf{R}^n$. Then the distribution of Y is a *multivariate normal distribution*, or *multivariate Gaussian distribution*, given with the multivariate density

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{n/2}|\mathbf{\Sigma}|^{1/2}} e^{-1/2(\mathbf{y}-\mu)'\mathbf{\Sigma}^{-1}(\mathbf{y}-\mu)}$$

where $\mathbf{\Sigma} = PP'$. This is denoted $Y \sim n(\mu, \mathbf{\Sigma})$ (or $Y \sim MVN(\mu, \mathbf{\Sigma})$).

Proof. Since $X_1, ..., X_n \sim n(0,1)$ iid, the joint density is given as the product

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{n} f_{\mathbf{X_i}}(\mathbf{x_i}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-x_i^2/2} = \frac{1}{(2\pi)^n/2} e^{-\sum x_i^2/2}.$$

The inverse of the function $\mathbf{x} \to \mathbf{y} = P\mathbf{x} + \mu$ is $\mathbf{y} \to \mathbf{x} = P^{-1}(\mathbf{y} - \mu) = g(\mathbf{y})$ with Jacobian determinant $J = |\frac{\delta g}{\delta y}| = |P^{-1}|$ so the density of $\mathbf{Y}$ is

$$f(y) = f_X(g(y))|J| = f_X(P^{-1}(\mathbf{y} - \mu))|P^{-1}|.$$

Since $\mathbf{\Sigma} = |PP'| = |P|^2 > 0$ we see that

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{n/2}|P|} e^{-[P^{-1}(\mathbf{y}-\mu)]'[P^{-1}(\mathbf{y}-\mu)]}$$

$$\Rightarrow f(\mathbf{y}) = \frac{1}{(2\pi)^{n/2}|\mathbf{\Sigma}|^{1/2}} e^{-(\mathbf{y}-\mu)'\mathbf{\Sigma}^{-1}(\mathbf{y}-\mu)}$$

(since $(P^{-1})'P^{-1} = (P')^{-1}P^{-1} = (PP')^{-1} = \mathbf{\Sigma}^{-1}$) - and in particular, this is in fact a density). ☐

Remark 5.1. Some comments

- The univariate normal is a special case
- If $\mathbf{\Sigma}$ is diagonal (i.e. $\text{Cov}(Y_i, Y_j) = 0$ if $i \neq j$), then the random variables are independent.

Theorem 5.3 If $X \sim n(\mu, \mathbf{\Sigma})$, then X_i, X_j are independent if and only if $\mathrm{Cov}(X_i, X_j) = 0$.

Theorem 5.4 If $(X_1,\ X_2, \ldots X_n,\ Y_1, Y_2, \ldots, Y_m)'$ is a Gaussian r.v., then $\mathbf{X} = (X_1,\ X_2, \ldots X_n)'$ and $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_m)'$ are independent iff $\mathrm{Cov}(X_i, Y_j) = 0\ \forall i, j$.

Theorem 5.5 Let $X_i \sim n(\mu, \sigma^2)$ be independent, $i = 1, ..., n$, and $Y_i := \boldsymbol{\xi_i'}\mathbf{X}$ where $\boldsymbol{\xi_1}, ..., \boldsymbol{\xi_n}$ form an orthonormal basis for $\mathbb{R}^n$. Then $Y_1, ..., Y_n$ are independent Gaussian random variables with

$$Y_i \sim n(\boldsymbol{xi_i}'\boldsymbol{\mu}, \sigma^2).$$

Proof. All of this follows from the definition of a multivariate normal distribution. □

Remark 5.2. The properties of the common t-test now follow from a collection of results based on the above. First let

$$\boldsymbol{\xi_1} := \frac{1}{\sqrt{n}}\mathbf{1},\ V := Span\{\boldsymbol{\xi_1}\}$$

and expand this (using e.g. a Gram-Schmidt process) to obtain $\boldsymbol{\xi_2}, ..., \boldsymbol{\xi_n}$ which form an orthonormal basis for $V^{\perp}$. Thus $\boldsymbol{\xi_1}, ..., \boldsymbol{\xi_n}$ form an orthonormal basis for $\mathbb{R}^n$. Write $X = \sum_{i=1}^{n} \hat{\zeta}_i \cdot \boldsymbol{\xi_i}$ - the coordinates of $\mathbf{X}$ in the basis (ξ_i) are $\hat{\zeta}_1, ..., \hat{\zeta}_n$ where $\hat{\zeta}_i = \mathbf{X} \cdot \xi_{\mathbf{i}}$ so that

1. $\xi_i = \frac{1}{\sqrt{n}} \sum_i X_i = \sqrt{n} \cdot X$ and $\sum_{i=2}^{n} \hat{\zeta}_i \boldsymbol{\xi_i} = \mathbf{X} - \hat{\zeta} \cdot \boldsymbol{\xi} = \mathbf{X} - \sqrt{n} \cdot \bar{X} \frac{1}{\sqrt{n}} \mathbf{1} = \mathbf{X} - \bar{X}\mathbf{1}$.
2. $\mathrm{Cov}(\hat{\zeta}_i, \hat{\zeta}_j) = 0$ if $i \neq j$.
3. $(\hat{\zeta}_1, ..., \hat{\zeta}_n)' = P\mathbf{X} \sim n(P\mu, \sigma^2 PP')$ with $P = [\xi_1' ... \xi_n']'$ and $PP' = I$.
4. $E\hat{\zeta}_i = \mathbb{E}[\mathbf{X} \cdot \xi_{\mathbf{i}}] = (\mu \cdot \mathbf{1}) \cdot \boldsymbol{\xi_i} = 0$ if $i \geq 2$
5. $\sum_{i=1}^{n} (X_i - \bar{X}) = \| \mathbf{X} - \bar{X} \|^2 = \| \sum \hat{\zeta}_i \cdot \xi_i \|^2 = \sum_{i=2} \hat{\zeta}_i^2$
6. For $i \geq 2$ we see that $\hat{\zeta}_i \sim n(0, \sigma^2)$ and these are independent so $\frac{\hat{\zeta}_i}{\sigma} \sim n(0, 1)$ are also independent
7. $\frac{\sum_{i=2}^{n} \hat{\zeta}_i^2}{\sigma^2} \sim \chi^2_{n-1}$ and independent of $\hat{\zeta}_1 \sim n(\sqrt{n}\mu, \sigma^2)$ and we obtain

$$\left.\begin{array}{l} \frac{\sum (X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n-1} \\ \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim n(0, 1) \end{array}\right\} \text{independent}$$

thus

$$\frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2/\sigma^2}{n-1}}} \sim t_{n-1}$$

Remark 5.3. Note that if $X_1, \ldots, X_n \sim n(\mu, \sigma^2)$ iid, then $\mathrm{E}\bar{X} = \mu$ and $\mathrm{E}S^2 = \sigma^2$, where $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i, S^2 = \frac{1}{n-1}\sum_{i=1}^n X_i - \bar{X})^2$. But we also see that e.g.

$$\mathrm{E}\bar{X} = \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^n X_i\right] = \frac{1}{n}\sum_{i=1}^n \mathrm{E}X_i = \frac{1}{n}n\mu = \mu,$$

which holds independently of any assumptions of normality - and the r.v.s do not have to be independent, *i.e.*: If $X_1, \ldots, X_n$ are random variables with $\mathrm{E}X_i = \mu$, then $\mathrm{E}\bar{X} = \mu$.

Remark 5.4. Next note that if $X_1, \ldots, X_n$ are independent random variables with expected value μ variance σ^2, then[1]:

$$\begin{aligned}
\mathrm{E}\left[\sum_{i=1}^n (x_i - \bar{X})^2\right] &= \mathrm{E}\left[\sum_{i=1}^n X_i^2 - n\bar{X}^2\right] \\
&= \sum_{i=1}^n \mathrm{E}[X_i^2] - n\mathrm{E}[\bar{X}^2] \\
&= \sum_{i=1}^n (\sigma^2 + \mu^2) - n(\sigma_{\bar{X}}^2 + \mu_{\bar{X}}^2) \\
&= n\sigma^2 + n\mu^2 - n\frac{\sigma^2}{n} - n\mu^2 \\
&= (n-1)\sigma^2.
\end{aligned}$$

We have shown: If $X_1, \ldots, X_n$ are independent with $\mathrm{E}X_i = \mu, \mathrm{V}X_i = \sigma^2$, then $\mathrm{E}\bar{X} = \mu$ and $\mathrm{E}S^2 = \sigma^2$.

[1] Where we use $\sigma_{\bar{X}}^2 = \sigma^2/n$ if the X_i are independent and a general formula: $\sigma^2 = \mathrm{E}[X^2] - \mu^2$, inverted to give the very useful version, $\mathrm{E}[X^2] = \sigma^2 + \mu^2$ for a random variable with this expected value and variance.