Chapter 2. Multivariate Distributions
2.3. Conditional Distributions and Expectations—Proofs of Theorems
Table of contents

1 Theorem 2.3.1
Theorem 2.3.1

**Theorem 2.3.1.** Let \((X_1, X_2)\) be a random vector such that the variance of \(X_2\) is finite. Then

(a) \(E[E[X_2 \mid X_1]] = E[X_2]\), and

(b) \(\text{Var}(E[X_2 \mid X_1]) \leq \text{Var}(X_2)\).

**Proof.** We give proofs for the continuous case and leave the discrete case as an exercise.

(a) We have

\[
E[X_2] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_2 f(x_1, x_2) \, dx_2 \, dx_1
\]

\[
= \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} x_2 \frac{f(x_1, x_2)}{f_1(x_1)} \, dx_2 \right) f_1(x_2) \, dx_1
\]

\[
= \int_{-\infty}^{\infty} E[X_2 \mid x_1] f_1(x_1) \, dx_1 = E[E[X_2 \mid X_1]]
\]

(notice that \(E[X_2 \mid x_1]\) is a function of \(x_1\)).
**Theorem 2.3.1.** Let \((X_1, X_2)\) be a random vector such that the variance of \(X_2\) is finite. Then

(a) \(E[E[X_2 | X_1]] = E[X_2]\), and

(b) \(\text{Var}(E[X_2 | X_1]) \leq \text{Var}(X_2)\).

**Proof.** We give proofs for the continuous case and leave the discrete case as an exercise.

**a)** We have

\[
E[X_2] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_2 f(x_1, x_2) \, dx_2 \, dx_1
\]

\[
= \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} x_2 \frac{f(x_1, x_2)}{f_1(x_1)} \, dx_2 \right) f_1(x_2) \, dx_1
\]

\[
= \int_{-\infty}^{\infty} E[X_2 | x_1] f_1(x_1) \, dx_1 = E[E[X_2 | X_1]]
\]

(notice that \(E[X_2 | x_1]\) is a function of \(x_1\)).
Proof (continued). (b) Let $\mu_2 = E[X_2]$, then

$$\text{Var}(X_2) = E[(X_2 - \mu_2)^2] = E[(X_2 - E[X_2 \mid X_1] + E[X_2 \mid X_1] - \mu_2)^2]$$

$$= E[(X_2 - E[X_2 \mid X_1] + E[X_2 \mid X_1] - \mu_2)(X_2 - E[X_2 \mid X_1] + E[X_2 \mid X_1] - \mu_2)]$$

$$= E[(X_2 - E[X_2 \mid X_1])^2 + (E[X_2 \mid X_1] - \mu_2)(X_2 - E[X_2 \mid X_1])]$$

$$+ (X_2 - E[X_2 \mid X_1])(E[X_2 \mid X_1] - \mu_2) + (E[X_2 \mid X_1] - \mu_2)^2$$

$$= E[(X_2 - E[X_2 \mid X_1])^2] + 2E[(E[X_2 \mid X_1] - \mu_2)(X_2 - E[X_2 \mid X_1])]$$

$$+ E[(E[X_2 \mid X_1] - \mu_2)^2]$$

where the last equality holds since $E$ is linear by Theorem 2.1.1.
Theorem 2.3.1 (continued 2)

Proof (continued).

\[
2E[(E[X_2 \mid X]1) - \mu_2)(X_2 - E[X_2 \mid X_1])] \\
= 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (X_2 - E[X_2 \mid x_1])(E[X_2 \mid x_1] - \mu_2)f(x_1, x_2) \, dx_2 \, dx_1 \\
= 2 \int_{-\infty}^{\infty} (E[X_2] - \mu_2) \left( \int_{-\infty}^{\infty} (x_2 - E[X_2 \mid x_1]) \frac{f(x_1, x_2)}{f_1(x_1)} \, dx_2 \right) f_1(x_1) \, dx_1.
\]

We have

\[
\int_{-\infty}^{\infty} (x_2 - E[X_2 \mid x_1]) \frac{f(x_1, x_2)}{f_1(x_1)} \, dx_2 \\
= \int_{-\infty}^{\infty} x_2 \frac{f(x_1, x_2)}{f_1(x_1)} \, dx_2 - E[X_2 \mid x_1] \int_{-\infty}^{\infty} \frac{f(x_1, x_2)}{f_1(x_1)} \, dx_2 \\
= E[X_2 \mid x_1] - EpX_2 \mid x_2](1) = 0,
\]

\ldots
Theorem 2.3.1 (continued)

Proof (continued). . . so
\[ 2E[(X_2 - E[X_2 | X_1])(E[X_2 | X_1] - \mu_2)] = 0 \]
and
\[ \text{Var}(X_2) = E[(X_2 - E[X_2 | X_1])^2] + E[(E[X_2 | X_1] - \mu_2)^2] \]
\[ \geq E[(E[X_2 | X_1] - \mu_2)^2] \]  \hspace{1cm} (*)
since \( E[(X_2 - E[X_2 | X_1])^2] \geq 0 \).

Now for random variable \( X \),
\[ \text{Var}(X) = E[(X - \mu)^2] = E[(X - E[X])^2] \] (see Definition 1.9.2), so
random variable \( E[X_2 | X_1] \) has mean \( E[E[X_2 | X_1]] \) and by part (a),
\( E[E[X_2 | X_1]] = E[X_2] = \mu_2 \) so that
\[ \text{Var}(E[X_2 | X_1]) = E[(E[X_2 | X_1] - \mu_2)^2] \]
and hence by (*)
\[ \text{Var}(X_2) \geq E[(E[X_2 | X_1] - \mu_2)^2] = \text{Var}(E[X_2 | X_1]), \]
as claimed.
Theorem 2.3.1 (continued 3)

Proof (continued). . . so $2E[(X_2 - E[X_2 \mid X_1])(E[X_2 \mid X_1] - \mu_2)] = 0$ and

$$\text{Var}(X_2) = E[(X_2 - E[X_2 \mid X_1])^2] + E[(E[X_2 \mid X_1] - \mu_2)^2]$$

$$\geq E[(E[X_2 \mid X_1] - \mu_2)^2] \quad (*)$$

since $E[(X_2 - E[X_2 \mid X_1])^2] \geq 0$. Now for random variable $X$, $\text{Var}(X) = E[(X - \mu)^2] = E[(X - E[X])^2]$ (see Definition 1.9.2), so random variable $E[X_2 \mid X_1]$ has mean $E[E[X_2 \mid X_1]]$ and by part (a), $E[E[X_2 \mid X_1]] = E[X_2] = \mu_2$ so that

$$\text{Var}(E[X_2 \mid X_1]) = E[(E[X_2 \mid X_1] - \mu_2)^2]$$

and hence by ($*$)

$$\text{Var}(X_2) \geq E[(E[X_2 \mid X_1] - \mu_2)^2] = \text{Var}(E[X_2 \mid X_1]),$$

as claimed.