Lecture 3 The Expectation Operator

Definition 1. Let $\vec{X} = (X_1, X_2, \cdots, X_n)$ be an n-D random variable with sample space S_X and with *pdf* $f_{\vec{X}}(\vec{x})$. Let $g(\vec{x})$ be any function of $\vec{x} = (x_1, x_2, \cdots, x_n)$. The **expectation operator E(*)** is defined as

$$\mu_{g(\mathbf{X})} \overset{\Delta}{=} E[g(\mathbf{X})] = \int_{S_\mathbf{X}} g(\mathbf{x})\, f_\mathbf{X}(\mathbf{x})\, d\mathbf{x} \tag{1}$$

[c.f. (1.7.5-6) on p.16 in the book.]

The integral (1) is an n-D integral (or sum, if $S_\mathbf{X}$ is discrete). It could be argued that (1) is the only equation one might need in order to understand anything and everything related to expected values.

1. Expected Values of Functions of 1-D *X*

***Example* 1.1** Consider the following examples of the function $g(*)$.

(i) $g(x) = x$: $E(X) = \int_{S_X} x\, f_X(x)\, dx \overset{\Delta}{=} \mu_X$. This is called the ***mean value*** for X.

(ii) $g(x) = x^2$: $E(X^2) = \int_{S_X} x^2\, f_X(x)\, dx \overset{\Delta}{=} \mu_{X^2}$. This is called the ***mean squared value*** for X.

(iii) $g(x) = (x - \mu_X)^2$: $E[(X-\mu_X)^2] = \int_{S_X} (x-\mu_X)^2\, f_X(x)\, dx \overset{\Delta}{=} \sigma_X^2$. This is called the ***variance*** of X.

(iv) $g(x) = e^{i\omega x}$: $E(e^{i\omega X}) = \int_{S_X} e^{i\omega x}\, f_X(x)\, dx \overset{\Delta}{=} \Phi_X(\omega)$. This is called the ***characteristic function*** for X.

(v) $g(x) = ax^2 + bx + c$: $E(aX^2 + bX + c) = \int_{S_X} (ax^2 + bx + c)\, f_X(x)\, dx$. This is just a quadratic function of X.

Now, let's express (v) in terms of the parameters in (i) and (ii):

$$E(aX^2 + bX + c) = a\int_{S_X} x^2 f_X(x)\, dx + b\int_{S_X} x f_X(x)\, dx + c\int_{S_X} f_X(x)\, dx .$$

Hence, we obtain : $E(aX^2 + bX + c) = a\,\mu_{X^2} + b\,\mu_X + c$. □

The above examples all follow the same procedure; namely, to compute $E[g(X)]$, you simply integrate $g(x)$ against $f_X(x)$.

Result 1.1 $\sigma_X^2 = \mu_{X^2} - \mu_X^2$. *Proof*: $\sigma_X^2 \overset{\Delta}{=} E[(X-\mu_X)^2] = E(X^2 - 2\mu_X X + \mu_X^2) = \mu_{X^2} - 2\mu_X^2 + \mu_X^2 = \mu_{X^2} - \mu_X^2$.

***Example* 1.2** Suppose. Let $Y = aX + b$.

Show that $\mu_Y = \mu_X + b$, and that $\sigma_Y = |a|\sigma_X$.

<u>Solution</u>: $\mu_Y = E(Y) = E(aX + b) = aE(X) + b = a\mu_X + b$.

$$\sigma_Y^2 \overset{\Delta}{=} E[(Y-\mu_Y)^2] = E[(aX + b - a\mu_X - b)^2] = a^2 E[(X-\mu_X)^2] \overset{\Delta}{=} a^2\sigma_X^2$$

Example 1.3 Let $X \sim Ber(p)$. Compute its *mean*, μ_X , it *mean-square*, μ_{X^2}, and its *variance*, σ_X^2 .

Solution: $\mu_X \overset{\Delta}{=} E(X) = \int_{S_X} x f_X(x)\,dx = \sum_{x=0}^{1} x \Pr[X = x] = 0(1-p) + 1(p) = p$.

$$\mu_{X^2} \overset{\Delta}{=} E(X^2) = \int_{S_X} x^2 f_X(x)\,dx = \sum_{x=0}^{1} x^2 \Pr[X = x] = 0^2(1-p) + 1^2(p) = p\,.$$

$$\sigma_X^2 = \mu_{X^2} - \mu_X^2 = p - p^2 = p(1-p)\,.$$

Example 1.4 Let $X \sim \exp(\lambda)$. Compute its *mean*, μ_X , it *mean-square*, μ_{X^2}, and its *variance*, σ_X^2 .

Solution: $\mu_X \overset{\Delta}{=} E(X) = \int_{S_X} x f_X(x)\,dx = \int_0^\infty x(\lambda e^{-\lambda x})\,dx\,.$ (1.1)

From a table of integrals, we have: $\int x e^{cx} dx = \frac{e^{cx}}{c^2}(cx-1)$. Applying this to (1.1) gives:

$$\mu_X = \lambda \int_0^\infty x e^{-\lambda x}\,dx = \lambda \left(\frac{e^{-\lambda x}}{\lambda^2}\right)(-\lambda x - 1)\bigg|_{x=0}^{\infty} = 0 - \left(\frac{-1}{\lambda}\right) = \frac{1}{\lambda}\,.$$

[Note: This is not a calculus course. We will use tables whenever possible. ☺].

Similarly, from a table of integrals, we have $\int x^2 e^{cx} dx = e^{cx}\left(\frac{x}{c} - \frac{2x}{c^2} + \frac{2}{c^3}\right)$. Hence,

$$\mu_{X^2} = \lambda \int_0^\infty x^2 e^{-\lambda x}\,dx = \lambda e^{-\lambda x}\left(\frac{x}{-\lambda} - \frac{2x}{\lambda^2} + \frac{2}{-\lambda^3}\right)\bigg|_{x=0}^{\infty} = \lambda\left[0 - \left(\frac{-2}{\lambda^3}\right)\right] = \frac{2}{\lambda^2}\,.$$

Finally, $\sigma_X^2 = \mu_{X^2} - \mu_X^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}\,.$

***Example* 1.5** For $X \sim N(\mu_X = 100, \sigma_X = 10)$, find the mean and standard deviation of $Y = 20\log_{10}(X)$ [e.g. converting sound pressure in bars to units of decibels].

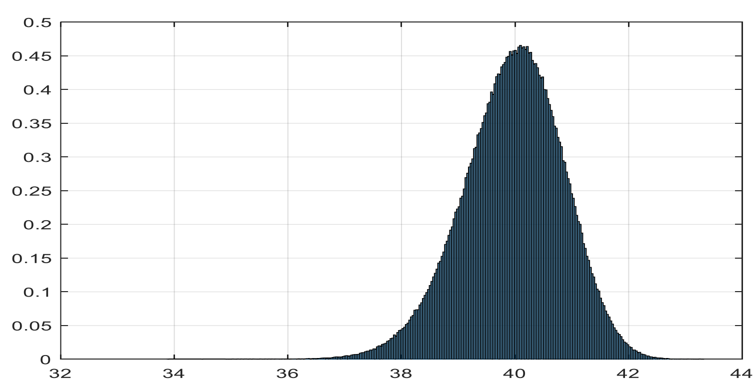
Solution: We could try to use a table of integrals to compute $\mu_Y \overset{\Delta}{=} E(Y) = E\{20\log_{10}(X)\} = 20\int_{S_X} \log(x) e^{-(x-\mu_X)^2/2\sigma_X^2}\,dx\,.$

It could be rewarding to find the closed form for such an integral. However, we would the need to find a similar integral to compute σ_Y . Here, we will simply use simulations

x=normrnd(100,10,10^6,1); y=20*log10(x);

muY=mean(y) = 39.9568 stdY=std(y) = 0.8795

Are these exact? No. Use 10^{10} simulations for greater accuracy. While not asked for, we also get the pdf at right 'for free'



>> histogram(y,'Normalization','pdf')

We see that it is a skewed pdf; one that might be well-modeled as a gamma pdf. □

2. Expected Values of Functions of 2-D $X=(X_1, X_2)$

Consider the 2-D random variable $\mathbf{X}=(X_1,X_2)$.

Definition 2.1 The *mean* of $\mathbf{X}=(X_1,X_2)$ is defined as $\boldsymbol{\mu}_{\mathbf{X}} \overset{\Delta}{=} (\mu_{X_1},\mu_{X_2})$.

***Example* 2.1** Consider the following examples of the function $g(*)$.

(i) $g(\mathbf{x})=\mathbf{x}=(x_1,x_2)$: $\boldsymbol{\mu}_{\mathbf{X}}=E(\mathbf{X})\overset{\Delta}{=}(\mu_{X_1},\mu_{X_2})$.

(ii) $g(\mathbf{x})=x_1$: $E[g(\mathbf{X})]=E[X_1]=\int\limits_{S_{\mathbf{X}}} x_1 f_{\mathbf{X}}(x_1,x_2)dx_2dx_1=\int\limits_{S_{X_1}} x_1[\int\limits_{S_{X2}} f_{\mathbf{X}}(x_1,x_2)dx_2]dx_1=\int\limits_{S_{X_1}} x_1 f_{X_1}(x_1)dx_1=E(X_1)$.

Note: This result relies directly on the fact that $f_{X_1}(x_1)=\int\limits_{S_{X2}} f_{\mathbf{X}}(x_1,x_2)dx_2$. It is used to show the next result.

(iii) $g(\mathbf{x})=ax_1+bx_2$:

$$E(aX_1+bX_2)=\int\limits_{S_{\mathbf{X}}}(ax_1+bx_2)f_{\mathbf{X}}(x_1,x_2)dx_2dx_1=a\int\limits_{S_{X_1}} x_1[\int\limits_{S_{X2}} f_{\mathbf{X}}(x_1,x_2)dx_2]dx_1+b\int\limits_{S_{X_2}} x_2[\int\limits_{S_{X_1}} f_{\mathbf{X}}(x_1,x_2)dx_1]dx_2=aE(X_1)+bE(X_2)$$

NOTE: This is a VERY important result. In words, it say that E(*) is a *linear operation.* In particular, the expected value of a sum is <u>always</u> the sum of the expected values.

***Property* 1: *E*(*aX* + *bY* + *c*) = *aE*(*X*) + *bE*(*Y*) + *c*. In words, this property states that *E*(*) is a *linear operator*.**

(iv) $g(\mathbf{x})=x_1x_2$ $E[g(\mathbf{X})]=E(X_1X_2)=\int\limits_{S_X} x_1x_2 f_{(X_1X_2)}(x_1,x_2)\,dx_1dx_2$.

(v) $g(\mathbf{x})=(x_1-\mu_{X_1})(x_2-\mu_{X_2})$: $E[g(\mathbf{x})]=E[(X_1-\mu_{X_1})(X_2-\mu_{X_2})]\overset{\Delta}{=}\sigma_{X_1X_2}$. This is the ***covariance*** between X_1 and X_2. We will, at times, also express this as $Cov(X_1,X_2)\overset{\Delta}{=}\sigma_{X_1X_2}$. From (iii) and (v) we also have: $E(X_1X_2)=\sigma_{X_1X_2}+\mu_{X_1}\mu_{X_2}$

(vi) We now present a second important property:

***Property* 2: $\mathbf{Var(aX+bY+c)=a^2Var(X)+b^2Var(Y)+2abCov(X,Y)}$**

[For the interested student, the proof of this property is given in the Appendix.]

***Special Case 1* (*b*=0):** Var(aX + c) = a^2Var(X)

***Special Case 2* (σ_{XY} = 0):** Var(aX + bY) = a^2Var(X) + b^2*Var*(Y)

***Special Case 3* (*a*=*b*=1 & σ_{XY} = 0):** Var(X + Y) = Var(X) + *Var*(Y)

Definition 2.2 The ***correlation coefficient*** between *X* and *Y* is defined as: $\rho_{XY} \overset{\Delta}{=} \dfrac{\sigma_{XY}}{\sigma_X \sigma_Y}$.

Definition 2.3 The matrix $Cov(\mathbf{X}) \overset{\Delta}{=} \begin{bmatrix} \sigma_{X_1}^2 & \sigma_{X_1X_2} \\ \sigma_{X_1X_2} & \sigma_{X_2}^2 \end{bmatrix} \overset{\Delta}{=} \Sigma_{\mathbf{X}}$ is called the ***covariance matrix*** associated with $\mathbf{X} = (X_1, X_2)$.

Definition 2.4 Random variables X_1 and X_2 are said to be ***uncorrelated*** if: $E(X_1 X_2) = E(X_1)E(X_2)$. (2.1)

They are said to be ***independent*** if: $f_{(X_1X_2)}(x_1, x_2) = f_{X_1}(x_1) \bullet f_{X_2}(x_2)$. (2.2)

Remark Any thoughtful student might ask: Why isn't the definition of *uncorrelated* defined in terms of the *correlation coefficient*, which has the range $-1 \le \rho_{XY} \le 1$? Indeed, this is most reasonable and natural. Unfortunately, as is sometimes the case in statistics, reason does not prevail ☹. Even so, we note that

$$\rho_{XY} \overset{\Delta}{=} \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \overset{\Delta}{=} \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y} = \frac{E(XY - Y\mu_X - X\mu_Y + \mu_X\mu_Y)}{\sigma_X \sigma_Y} = \frac{E(XY) - E(Y)\mu_X - E(X)\mu_Y + \mu_X\mu_Y}{\sigma_X \sigma_Y} = \frac{E(XY) - \mu_X\mu_Y}{\sigma_X \sigma_Y}.$$

Hence, we will have $\rho_{XY} = 0$ if and only if $E(XY) = E(X)E(Y)$. Hence, should one choose, one could redefine the definition of *X* and *Y* being *uncorrelated* to mean that $\rho_{XY} = 0$.

Result 2.1 If random variables X_1 and X_2 are independent, then they are uncorrelated.

Proof:

$$E(X_1X_2) = \int_{S_{X_2}} \int_{S_{X_1}} x_1x_2 f_{(X_1X_2)}(x_1, x_2)\, dx_1dx_2 = \int_{S_{X_2}} \int_{S_{X_1}} x_1x_2 f_{X_1}(x_1) f_{X_2}(x_2)\, dx_1dx_2 \quad \textit{by independence}$$

$$= \int_{S_{X_2}} x_2 f_{X_2}(x_2)dx_2 \int_{S_{X_1}} x_1 f_{X_1}(x_1)dx_1 \quad \textit{by calculus} \quad = E(X_2)E(X_1) \quad \textit{by definition.}$$

Thus, we see that if two random variables are independent, then they are uncorrelated. However, the converse is ***not*** necessarily true. Uncorrelatedness only means that they are not related in a *linear* way. **This is important!** Many engineers assume that because *X* and *Y* are uncorrelated, they have nothing to do with each other (i.e. they are independent). It may well be that they are, in fact, very related to one another.

***Example* 2.2** Let (X, Y) denote the act of measuring the length and width of a rectangular membrane. Assume that *X* and *Y* are mutually independent. Compute the mean and standard deviation for the computed area $A = XY$.

Solution: Because of independence, $\mu_A = E(A) = E(XY) = E(X)E(Y) = \mu_X\mu_Y$.

To compute σ_A, first compute $\mu_{A^2} = E(X^2Y^2) = E(X^2)E(Y^2) = (\mu_X^2 + \sigma_X^2)(\mu_Y^2 + \sigma_Y^2) = (\mu_X\mu_Y)^2 + (\sigma_X^2\sigma_Y^2) + (\sigma_X^2\mu_Y^2 + \sigma_Y^2\mu_X^2)$.

Then $\sigma_A^2 = \mu_{A^2} - \mu_A^2 = (\sigma_X^2\sigma_Y^2) + (\sigma_X^2\mu_Y^2 + \sigma_Y^2\mu_X^2)$, or $\sigma_A = \sqrt{\sigma_X^2\sigma_Y^2 + (\sigma_X^2\mu_Y^2 + \sigma_Y^2\mu_X^2)}$. Notice that the units of this expression are those of area.

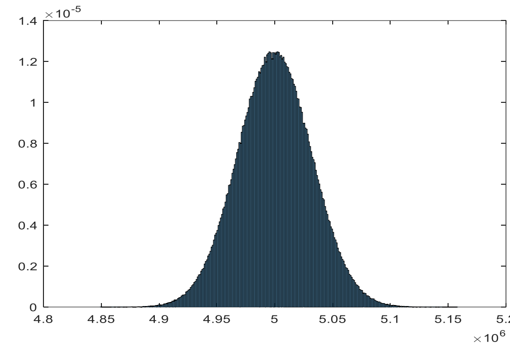
Suppose now that the area corresponds to that of a large field, with $X \sim N(1000\,m, 5m)$ and $Y \sim N(5000\,m, 20m)$. Then $\mu_A = \mu_X\mu_Y = 10^3(5\times10^3 = 5(10^6)\,m^2$. Furthermore, we can approximate σ_A as

$$\sigma_A \cong \sqrt{\sigma_X^2\mu_Y^2 + \sigma_Y^2\mu_X^2} = \mu_X\mu_Y\sqrt{(\sigma_X/\mu_X)^2 + (\sigma_Y/\mu_Y)^2} \overset{\Delta}{=} \mu_X\mu_Y\sqrt{r_X^2 + r_Y^2}$$

where $r_X = \sigma_X / \mu_X$ is a measure of the *relative uncertainty* in X. For the given numbers, we have $r_X = 5(10^{-3})$ and $r_Y = 4(10^{-3})$. Hence, $\sigma_A \cong (5\times10^6)(10^{-3})\sqrt{5^2+4^2} = 5\sqrt{41}(10^3) \cong 32(10^3)\ m^2$.

Writing $\sigma_A = \mu_X \mu_Y r_X \sqrt{1+(r_Y / r_X)^2}$ makes it clear as to how the various parameters contribute to this uncertainty.

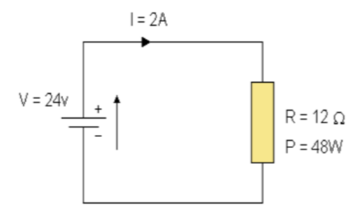
Finally, we offer the simulation-based *pdf* for *A*.



```
>> x=normrnd(1000,5,10^6,1); >> y=normrnd(5000,20,10^6,1);
>> a=x.*y;  >> histogram(a,'Normalization','pdf')
>> mean(a) = 5.0000e+06    std(a) = 3.2032e+04
```

Not only have we validated our theoretical results. We can also deduce that $A \overset{\bullet}{\sim} N[5(10^6)\,m^2\ , 32(10^2)\,m^2]$. With this, we can compute all manner of probabilities related to *A*. □

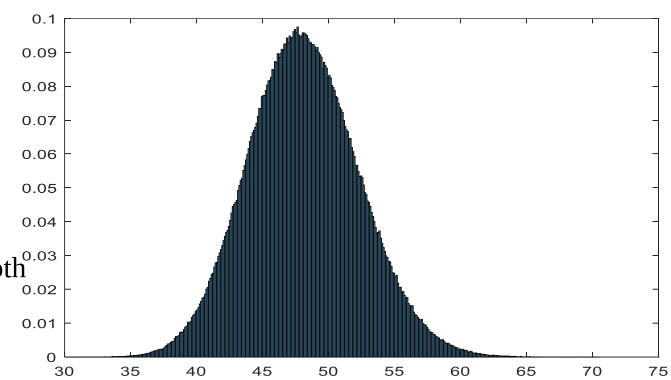
***Example* 2.3** Consider the design circuit shown at right. Assume $V \sim N(24v\,, 1v)$ is independent of $R \sim N(12\Omega\,, 0.3\Omega)$. The power draw is $P = V^2 / R$. Using a first order Taylor series expansion, we have $P \cong \mu_V^2 / \mu_R + \left(\frac{\partial P}{\partial V}\right)_{(\mu_V,\mu_R)}(V-\mu_V) \ + \left(\frac{\partial P}{\partial R}\right)_{(\mu_V,\mu_R)}(R-\mu_R)$ where



$\left(\frac{\partial P}{\partial V}\right)_{(\mu_V,\mu_R)} = \frac{2\mu_V}{\mu_R} = 4$ and $\left(\frac{\partial P}{\partial R}\right)_{(\mu_V,\mu_R)} = -\frac{\mu_V^2}{\mu_R^2} = -4$. So: $P \cong 48 + 4(V-\mu_V) - 4(R-\mu_R)$.

Hence: $\mu_P = E(P) \cong 48 + 4E(V-\mu_V) - 16E(R-\mu_R) = 48$ and $\sigma_P \cong \sqrt{16\sigma_V^2 + 16\sigma_R^2} \cong \sqrt{16+1.44} \cong 4.18$.

We will validate these approximations using simulations.



```
>> v=normrnd(24,1,10^6,1); >> r=normrnd(12,0.3,10^6,1);
>> p=v.^2 ./r; >> histogram(p,'Normalization','pdf')
 mean(p) = 48.1167      std(p) = 4.1847.
```

We see that our approximations $\mu_P \cong 48$ and $\sigma_P \cong 4.18$ are both slightly low. However, for a first order approximation, they're not too bad, either. We also see that *P* has a *normal pdf.* □

***Example* 2.4** Let (X,Y) denote the act of measuring the amount of time that a user of a hacked pc has wasted due to the hack, and the amount of time needed to fix the problem. Suppose that we are given the following information:

$$\mu_X = 7\,hrs. \quad \sigma_X = 1\,hr. \quad \mu_Y = 36\,hrs. \quad \sigma_Y = 15\,hrs. \quad \rho_{XY} = 0.7$$

Let $W = X + Y$ denote the total lost time associated with a hack. Find μ_W and σ_W.

Solution: $\mu_W = E(W) = E(X+Y) = E(X) + E(Y) = 7 + 36 = 43\,hrs.$

$\sigma_W^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY}$ where $\rho_{XY} = 0.7 = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} = \frac{\sigma_{XY}}{1(15)}$. Hence, $\sigma_{XY} = 0.7(15) = 10.5\,hrs^2$.

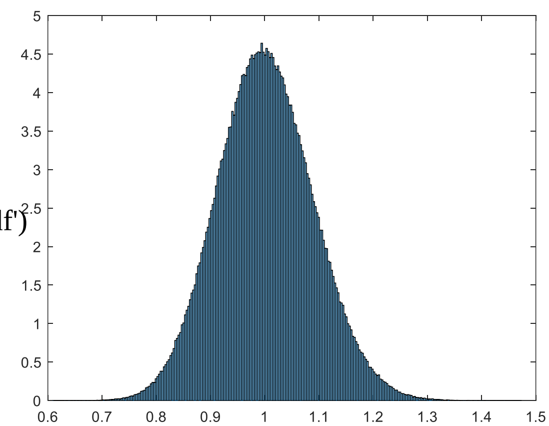
And so: $\sigma_W^2 = 1^2 + 15^2 + 2(10.5) = 247\,hrs^2$, hence $\sigma_W = 15.7\,hrs.$

Remark. Had we blindly used the formula $\sigma_W^2 = \sigma_X^2 + \sigma_Y^2$, which ignores the correlation between *X* and *Y*, we would have obtained $\sigma_W = 15\,hrs$. □

***Example* 2.5** Let $\mathbf{X} = (X_1, X_2)$ denotes the act of measuring the peak pressure and temperature of a given reaction. Suppose that the pressure is $X_1 \sim N(2000{,}100)$ and the temperature is $X_2 \sim N(5000{,}200)$. If pressure and temperature are correlated with $\rho = 0.9$, then $\sigma_{XY} = \rho\sigma_X\sigma_Y = 0.9(100)(200) = 1.8(10^4)$. Suppose that we have defined the following performance metric for the reaction: $P = (X_1X_2)/(\mu_{X_1}\mu_{X_2})$.

Compute μ_P: From (vi) in *Example* 1 we have $\mu_P = \mu_{X_1}\mu_{X_2} = E(X_1X_2) = \sigma_{X_1X_2} + \mu_{X_1}\mu_{X_2} = 1.8(10^4) + 10^7 \cong 10^7$.

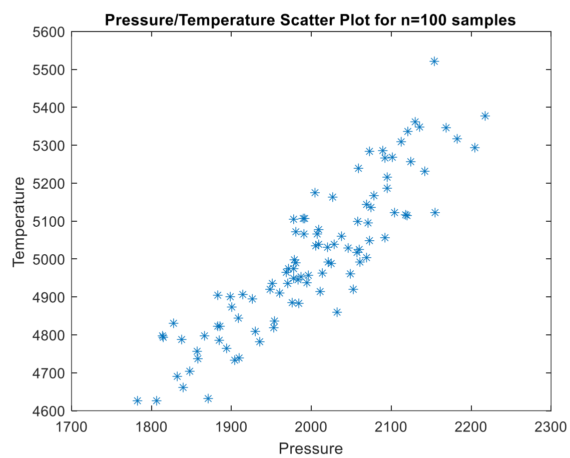
Hence, $\mu_P \cong 1.0$. We could compute σ_P in a similar fashion. Instead, we will use simulation.



```
>> muX=[2000,5000];   >> covX=[100^2 , 1.8e4 ; 1.8e4 , 200^2];
>> x=mvnrnd(muX,covX,1,10^6);   >> x=mvnrnd(muX,covX,10^6);
>> size(x)   ans =1000000          2
>> p=(x(:,1).*x(:,2))/(2000*5000);   >> histogram(p,'Normalization','pdf')
>> mean(p) =1.0018      >> std(p) = 0.0878
```

Conclusion: $P \overset{\bullet}{\sim} N(1.0\,,\,0.088)$.

Simulate a sample size of 100 measurements of $\mathbf{X} = (X_1, X_2)$ to arrive at a simulated scatter plot.



```
>> x=mvnrnd(muX,covX,100);
>> plot(x(:,1),x(:,2),'*')
>> xlabel('Pressure')
>> ylabel('Temperature')
>> title('Pressure/Temperature Scatter Plot for n=100 samples')
```

Estimate the correlation from the data.

```
>> R=corrcoef(x)
R =
    1.0000    0.8966
    0.8966    1.0000
```

The estimated correlation is the off-diagonal element 0.8966. □

There is one last important property that we will address in these notes:

***Property* 3: Cov(*aX*+*bY*+c, *W*) = *a*Cov(*X*,*W*) + *b*Cov(*Y*,*W*)**

[For the interested student, the proof of this property is given in the Appendix.]

***Example* 2.6** For the 2-D random variable (X,Y), consider the *Y-predictor*: $\widehat{Y} = mX + b$.

We will place two conditions on $\widehat{Y}$: (C1) $\mu_{\widehat{Y}} = \mu_Y$ and (C2) $Cov(Y - \widehat{Y}, X) = 0$. Find *m* and *b*.

Solution: From (P3): $Cov(Y - \widehat{Y}, X) = Cov(Y, X) - Cov(\widehat{Y}, X) = 0$. Also: $Cov(\widehat{Y}, X) = Cov(mX + b, X) = mCov(X, X)$.

These two equations give: $m = Cov(Y, X) / Cov(X, X) = \sigma_{XY} / \sigma_X^2$.

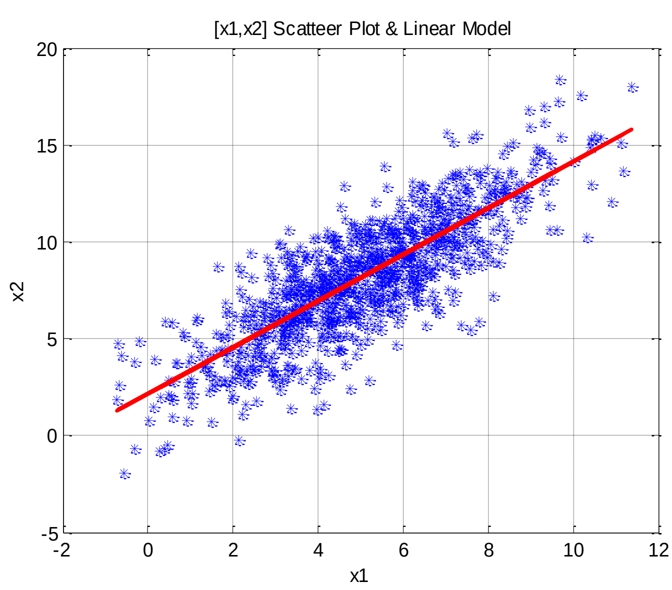
In relation to (C1), then: $E(mX + b) = m\mu_X + b = \mu_Y$. Hence: $b = \mu_Y - m\mu_X$.

***Example* 2.7** Recall the *Matlab* computations: For an $n \times 2$ data set $\mathbf{xy} \overset{\Delta}{=} (\mathbf{x}, \mathbf{y})$, the command 'mean(**xy**)' gives $[\widehat{\mu}_X \ , \ \widehat{\mu}_Y]$, and the command 'cov(**xy**)' gives $\begin{bmatrix} \widehat{\sigma}_X^2 & \widehat{\sigma}_{XY} \\ \widehat{\sigma}_{XY} & \widehat{\sigma}_Y^2 \end{bmatrix}$. Hence, only two commands are needed prior to computing the above estimates of *m* and *b*. Suppose that, in truth, we have:

$\mu_X = 5 \quad \sigma_X = 2 \quad \mu_Y = 8 \quad \sigma_Y = 3 \quad \rho_{XY} = 0.8$. In the code below, instead of using (X,Y) we will use $\mathbf{X} = (X_1, X_2)$ for personal convenience.

Matlab Demonstration:

```
% PROGRAM NAME: Ch5_2Ddata.m
% Matlab to simulate data from (X1,X2)
% Linear Model: X2hat = aX1 + b
% Parameter Values:
muX = [5 8]; % means
corr12 = 0.8; % correlation coefficient
var1 = 4; var2 = 9; % variances
cov12 = corr12*(var1*var2)^.5; % covariance
SIGMA = [var1 cov12 ; cov12 var2]; % cov.matrix
a = SIGMA(1,2)/SIGMA(1,1); % true slope
b = muX(2) - a*muX(1); % true intercept
%-----------------------------------------------
X = mvnrnd(muX,SIGMA,1000); % 1000 msmts of X
%-----------------------------------------------
% Estimate slope & intercept
muXhat = mean(X); % estimated means
SIGMAhat = cov(X); % estimated covariance matrix
ahat = SIGMAhat(1,2)/SIGMAhat(1,1); % est. slope
bhat = muXhat(2) - ahat*muXhat(1); % est. intercept
[a ahat ; b bhat]
%-----------------------------------------------
X2hat = ahat*X(:,1) + bhat; % X2 predictions
figure(1)
plot(X(:,1),X(:,2),'*')
hold on
plot(X(:,1),X2hat,'r','LineWidth',2)
grid
xlabel('x1')
ylabel('x2')
title('[x1,x2] Scatter Plot & Prediction')
```



[a ahat] = [1.2000 1.1988]

[b bhat] = [2.0000 2.1530]

□

Appendix

Proof of Property 2: $\sigma_W^2 = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\sigma_{XY}$

Proof: To simplify things, define $W \overset{\Delta}{=} aX + bY + c$. Recall that $\sigma_W^2 = Var(W) = E[(W-\mu_W)^2]$. Then

$Var(W) = E[(W-\mu_W)^2] = E[(aX + bY + c - a\mu_X - b\mu_Y - c)^2] = E\{[a(X-\mu_X) + b(Y-\mu_Y)]^2\}$.

Using the fact that $(u+v)^2 = u^2 + v^2 + 2uv$ gives

$Var(W) = E\{[a(X-\mu_X) + b(Y-\mu_Y)]^2\} = E[a^2(X-\mu_X)^2 + b^2(Y-\mu_Y)^2 + 2ab(X-\mu_X)(Y-\mu_Y)]$.

Finally, using linearity of E(*):

$\sigma_W^2 = a^2E[(X-\mu_X)^2] + b^2E[(Y-\mu_Y)^2] + 2abE[(X-\mu_X)(Y-\mu_Y)] = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\sigma_{XY}$. □

Proof of Property 3: $Cov(aX + bY + c\,, W) = aCov(X, W) + bCov(Y, W)$

Proof: By definition, we have: $Cov(aX + bY + c\,, W) = E\{[(aX + bY + c) - E(aX + bY + c)]\bullet[W - E(W)]\}$.

Using Property 1: $Cov(aX + bY + c\,, W) = E\{[(aX + bY + c) - (a\mu_X + b\mu_Y + c)]\bullet(W - \mu_W)\}$.

Regrouping terms: $Cov(aX + bY + c\,, W) = E\{[a(X-\mu_X) + b(Y-\mu_Y)]\bullet(W-\mu_W)\}$.

Again, from Property 1: $Cov(aX + bY + c\,, W) = aE\{a(X-\mu_X)(W-\mu_W)\} + bE\{(Y-\mu_Y)](W-\mu_W)\}$.

Hence, by definition: $Cov(aX + bY + c\,, W) = a\,\sigma_{XW} + b\sigma_{YW}$ □