Kom 505E week 8

22.11.2016

Covariance for std bivariate Gaussian pdf:

(Ex 12.13) Recall $X \sim \mathcal{N}(0,1)$
$Y \sim \mathcal{N}(0,1)$

$$\mathrm{Cov}(X,Y) = E_{X,Y}[XY]$$

$$= \iint_{-\infty}^{\infty} xy \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)} \cdot (x^2 - 2\rho xy + y^2)\right] dx\,dy$$

$$q = x^2 - 2\rho xy + y^2 + \rho^2 x^2 - \rho^2 x^2$$
$$= (y-\rho x)^2 + (1-\rho^2)x^2$$

$$\mathrm{Cov}(X,Y) = \iint_{-\infty}^{\infty} xy \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}(y-\rho x)^2\right] dx\,dy \quad \exp\left[-\frac{1}{2}x^2\right]$$

$$= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \underbrace{\int_{-\infty}^{\infty} y \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left[-\frac{(y-\rho x)^2}{2(1-\rho^2)}\right] dy}_{E_Y[Y] \Rightarrow Y \sim \mathcal{N}(\rho x, (1-\rho^2))} dx.$$

$$= \int_{-\infty}^{\infty} \rho x^2 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \rho \underbrace{E_X[X^2]}_{=1} \quad (X \sim \mathcal{N}(0,1))$$

$(\mathrm{Var} + \mu^2)$, $1 + 0^2$

$$\boxed{\mathrm{Cov}(X,Y) = \rho}$$

Recall: $\rho_{X,Y} = \dfrac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} = \rho$ (the correlation coeff.).

$\mathrm{Cov}(X,Y) = \rho$; $\mathrm{Var}(X) = 1$, $\mathrm{Var}(Y) = 1$ (written as "=ρ" and ditto marks)

★ If $\rho = 0$, and X & Y are jointly Gaussian distributed (e.g. std bivariate Gaussian), then

$$P_{X,Y}(x,y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} = P_X(x)\,P_Y(y)$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$$

then

X & Y are independent.

(Note. This also holds for general bivariate Gaussian pdfs $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$ & $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$.)

Note: !!! uncorrelated $\Rightarrow$ independent ONLY for Gaussian pdfs.
(Not true in general).

Def: Covariance Matrix $\underline{\underline{C}}$:

$$\underline{\underline{C}} = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix} = \begin{bmatrix} \mathrm{Var}X & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(X,Y) & \mathrm{Var}Y \end{bmatrix}$$

The general form bivariate Gaussian pdf:

$$P_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{\det C}} \exp\left(-\frac{1}{2} \begin{bmatrix} x-\mu_x \\ y-\mu_y \end{bmatrix}^T C^{-1} \begin{bmatrix} x-\mu_x \\ y-\mu_y \end{bmatrix}\right)$$

$$(\underline{x}-\underline{\mu})^T_{1\times 2}\, \underline{\underline{C}}^{-1}_{2\times 2} (\underline{x}-\underline{\mu})_{2\times 1}$$

★ If $\underline{\underline{C}}$ is diagonal, then X & Y are independent.

Theorem (12.7.1) Linear transformation of Gaussian r.v.s

Let $\begin{pmatrix} X \\ Y \end{pmatrix} \sim N(\underline{\mu}, \underline{\underline{C}})$, Let $\begin{pmatrix} W \\ Z \end{pmatrix} = \underline{\underline{G}} \begin{pmatrix} X \\ Y \end{pmatrix}$

then $\begin{pmatrix} W \\ Z \end{pmatrix} \sim N(\underline{\underline{G}}\,\underline{\mu}, \underline{\underline{G}}\,\underline{\underline{C}}\,\underline{\underline{G}}^T)$

Pf: $$p_{W,Z}(w,z) = p_{X,Y}\left(G^{-1}\begin{pmatrix} w \\ z \end{pmatrix}\right) \underbrace{\left|\det\left(\frac{\partial(x,y)}{\partial(w,z)}\right)\right|}_{|\det G^{-1}|}$$

$$= \frac{1}{2\pi\sqrt{\det C}\,|\det G|} \exp\left[-\frac{1}{2}\begin{pmatrix} w \\ z \end{pmatrix}^T \underbrace{G^{-T} C^{-1} G^{-1}}_{C_{new}^{-1}} \begin{pmatrix} w \\ z \end{pmatrix}\right]$$

✓ (exercise: go through this pf).

★ Transforming correlated Gaussian r.v.s to uncorrelated (= independent) Gaussians.

Let $\begin{pmatrix} X \\ Y \end{pmatrix} \sim N(\underline{\mu}, \underline{\underline{C}})$

Find a $\underline{\underline{G}}$ s.t. $\underline{\underline{G}} \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} W \\ Z \end{bmatrix}$ has a diagonal covariance.

$\equiv$ Need a $\underline{\underline{G}}$ s.t. $\boxed{\underline{\underline{G}}\,\underline{\underline{C}}\,\underline{\underline{G}}^T}$ is diagonal.

$$\underline{\underline{C}} = \underline{\underline{V}}\,\underline{\underline{\Lambda}}\,\underline{\underline{V}}^T = [v_1 \mid v_2] \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \end{bmatrix}$$

$$\Rightarrow \Lambda = \underline{\underline{V}}^T \underline{\underline{C}}\, \underline{\underline{V}}$$

$\Big($ $\underline{\underline{V}}$: orthogonal matrix of eigen vectors $\Big)$, $\underline{\underline{V}}^{-1} = \underline{\underline{V}}^T$

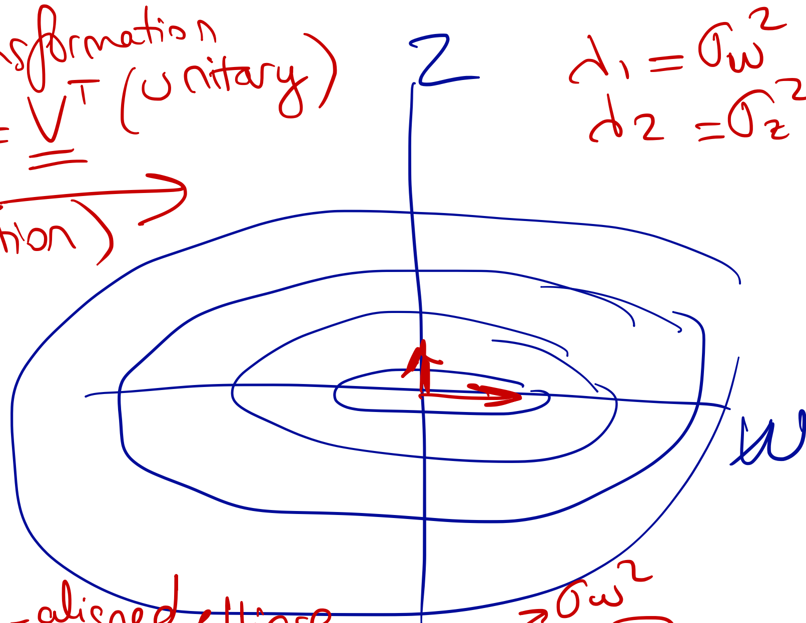
$$\Rightarrow \underline{\underline{G}} = \underline{\underline{V}}^T$$ $\Rightarrow$ resulting cov matrix is diagonal.

* Whitening transform: $\underline{G} = \Lambda^{-1/2}\, \underline{V}^T$ $\qquad \lambda_1 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

$$\underline{G}\,\underline{C}\,\underline{G}^T = \Lambda^{-1/2} \underbrace{V^T (V}_{I} \Lambda \underbrace{V^T) V}_{I} \Lambda^{-1/2} = \Lambda^{-1/2} \Lambda \Lambda^{-1/2} = I$$

angled ellipse
(correlated variables)

Transformation
$\underline{G} = \underline{V}^T$ (unitary)
(Rotation)

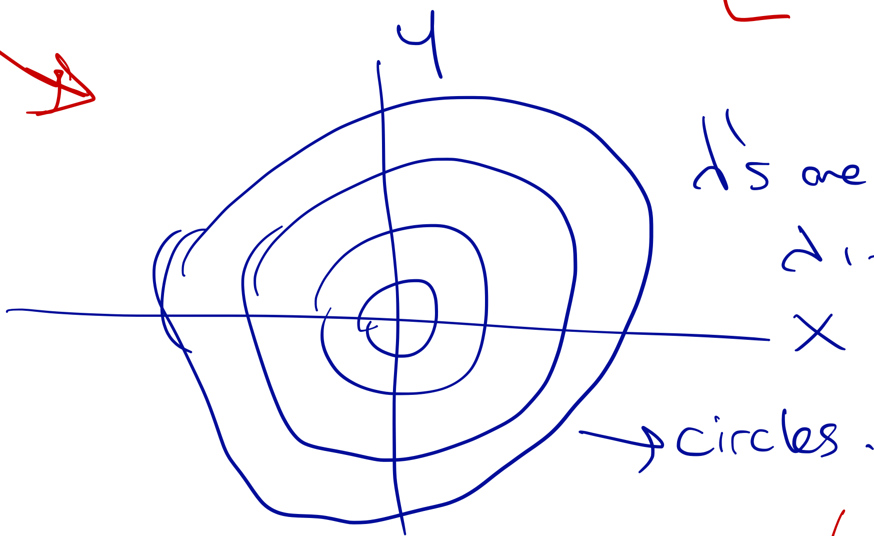


$\lambda_1 = \sigma_w^2$
$\lambda_2 = \sigma_z^2$

axis-aligned ellipse

$$C_{new} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \qquad \lambda_1 \to \sigma_w^2, \quad \lambda_2 \to \sigma_z^2$$

whitening



λ's are equal
$\lambda_1 = \lambda_2 = 1.$

circles.

Prediction of an R.V. from another $\begin{pmatrix} 12.9 \\ 7.9 \end{pmatrix}$

Assume X & Y jointly distrib & joint pdf is known.

* Covariance btw X & Y, relates to predictability of Y based on knowledge of X.

* like to estimate Y from X in a "linear" fashion:

$$\hat{Y} = g(X) = aX + b$$, find a & b for "best" prediction

best → lowest MSE.

$$MSE(a,b) = E_{X,Y}[(Y-\hat{Y})^2]$$
$$= E_{X,Y}[(Y-(aX+b))^2]$$

<u>Goal:</u> To minimize MSE w.r.t. a, b

$$\begin{cases} \frac{\partial}{\partial a} MSE(a,b) = \frac{\partial}{\partial a} E_{X,Y}[Y^2 + a^2X^2 - 2aXY + b^2 - 2(Y-aX)b] \\ \quad \rightarrow = 0 \\ \frac{\partial}{\partial b} MSE(a,b) = 0 \end{cases}$$

exercise:

→ solve the set of eqns to get "optimal" value for a & b.

$$a_{opt} = \frac{cov(X,Y)}{Var(X)}, \quad b_{opt} = \mu_Y - \frac{cov(X,Y)}{Var(X)}\mu_X$$

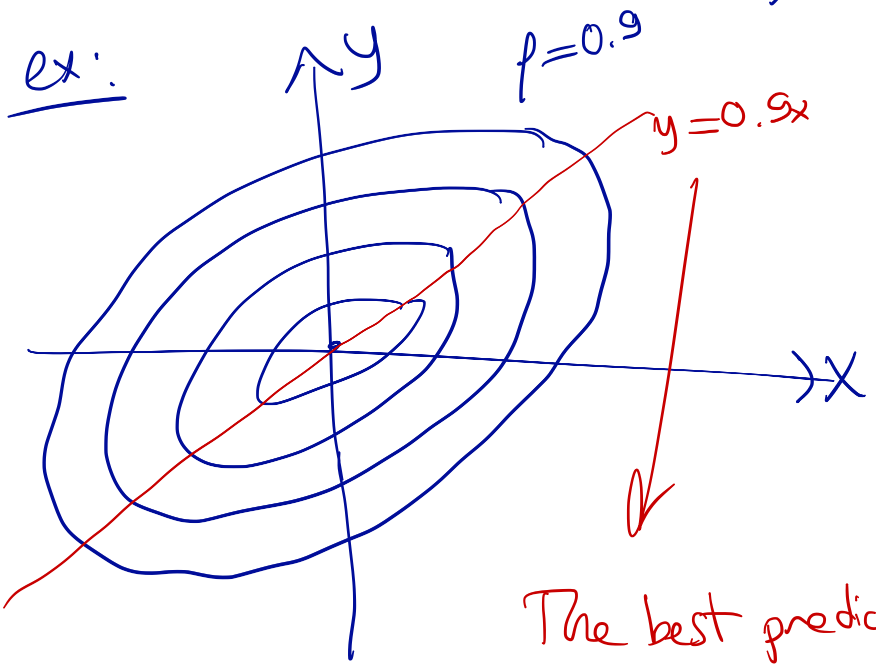
$$\Rightarrow \hat{Y} = a_{opt} X + b_{opt}$$
$$\hat{Y} = \mu_y + \frac{cov(X,Y)}{Var(X)}(X - \mu_x)$$

Note: If $Cov(X,Y) = 0 \Rightarrow \hat{Y} = \mu_y$ (X provides no info on Y)

for jointly Gaussian X, Y :

$$\hat{y} = \mu_y + \rho \frac{\cancel{\sigma_x} \sigma_y}{\sigma_x^{\cancel{2}}} (x - \mu_x)$$

ex:



$Var(X) = Var(Y) = 1$

$$\underline{\mu} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$C = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$$

$$\hat{y} = 0 + 0.9x$$

The best prediction of Y when $X = x$ is observed is given by a line.

12.8 Joint moments

: (Reading assignment: ~~Real world~~ 12.12 Optical Character Recog based on moments)

$E_{X,Y}[X^k Y^l]$: (k^{th}, l^{th})-order joint moment.

$$\triangleq \int \cdots \int x^k y^l \, p_{X,Y} \cdots dy$$

$E_{X,Y}[(X - \mu_X)^k (Y - \mu_Y)^l]$: centered (k^{th}, l^{th}) joint moment.

$\therefore Cov(X, Y) = E_{X,Y}[(X - \mu_X)(Y - \mu_Y)]$

$\hookrightarrow$ (1, 1)-order joint centralized moment of X, Y.

* If X & Y are indep: $E_{X,Y}[X^k Y^l] = E_X[X^k] E_Y[Y^l]$

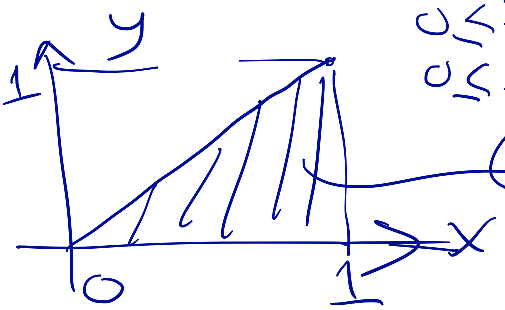
* If X & Y are indep r.v.s then so are $g(X)$ & $h(Y)$.

Ch 8 & 13 Conditional pdfs (pmfs):

$p(A|B) = \checkmark$ i.t.o events

Define $p_{Y|X}(y|X=x) = \dfrac{p_{X,Y}(x,y)}{p_X(x)}$ i.t.o pdfs

ex: * Joint pdf for (X,Y):

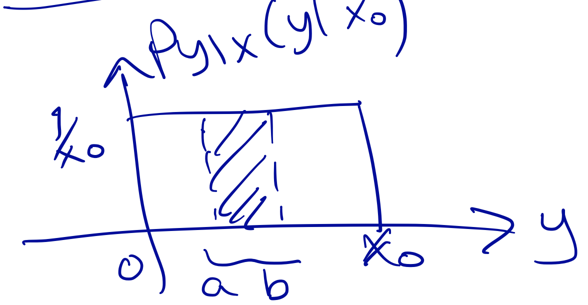


$0 \le x \le 1$
$0 \le y \le 1$
$y \le x$

$p_{X,Y}(x,y) = 2$

$$p_X(x) = \int_{-\infty}^{\infty} p_{X,Y}(x,y)\,dy = \int_0^x 2\,dy = 2x, \quad 0 \le x \le 1.$$

$$p_{Y|X}(y|x_0) = \frac{p_{X,Y}(x_0,y)}{p_X(x_0)} = \frac{2}{2x_0} = \begin{cases} \frac{1}{x_0}, & 0 \le y \le x_0 \\ 0, & \text{o/w.} \end{cases}$$



eg. $p(a \le Y \le b \mid X = x_0)$

$$= \int_a^b p_{Y|X}(y|x_0)\,dy$$

area $= (b-a)\dfrac{1}{x_0}$.

* $p_{Y|X}(y|x)$: probability per unit length of Y when $X=x$: $\left\{x - \frac{\Delta x}{2} \le X \le x + \frac{\Delta x}{2}\right\}$

family of pdfs with y as the independent variable. For each value of x, we have a different $p_{Y|X}$ (pdf).

this is what we think of when we condition on $X = x$.

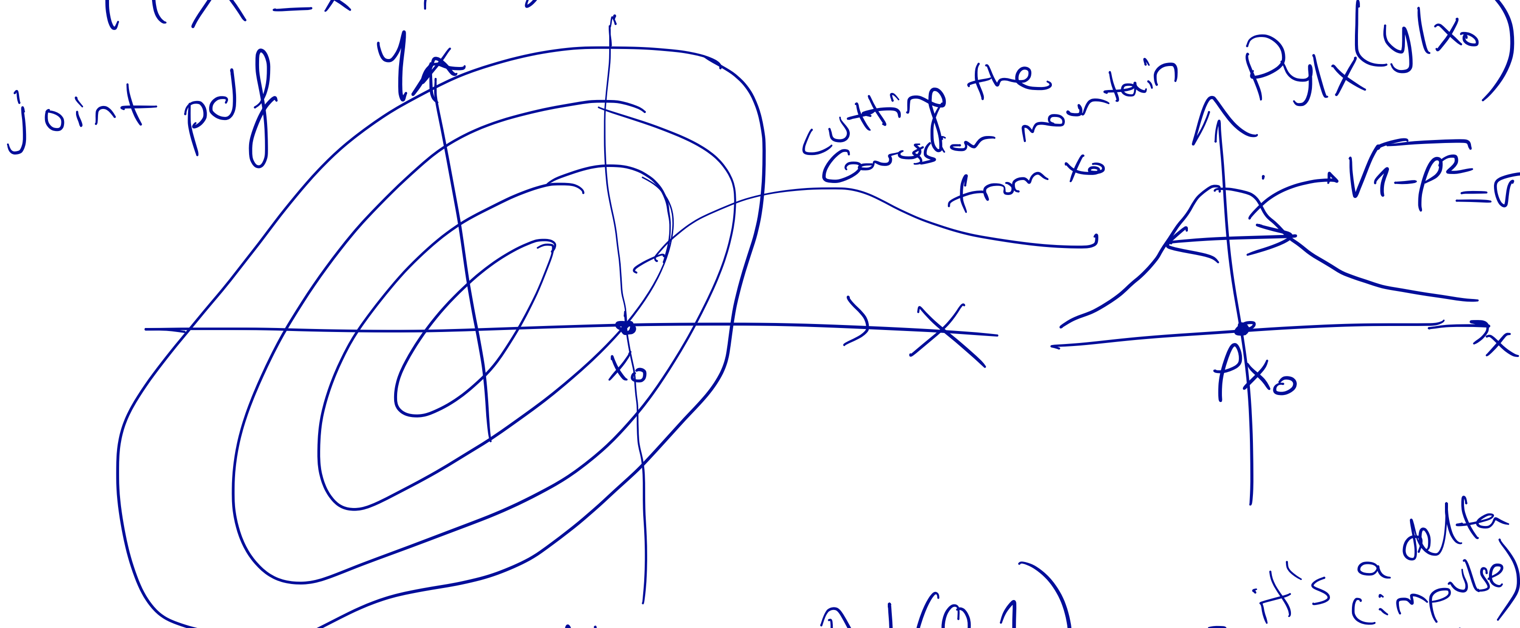
Ex: (X,Y) are s.b.n. $\mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}\right)$

$$P_{Y|X}(y|x) = \frac{P_{X,Y}(x,y)}{P_X(x)} = \frac{\frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right)}{\frac{1}{\sqrt{2\pi}} e^{-x^2/2}}$$

Go thru a similar derivation as in 1st hour. (Check Derivation)

$$= \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left(-\frac{1}{2(1-\rho^2)}(y-\rho x)^2\right)$$

$$Y | X=x \sim \mathcal{N}(\rho x, (1-\rho^2))$$



* if $\rho = 0$; $Y|X \sim \mathcal{N}(0,1)$

* if $\rho = 1$; $Y|X=x_0 \sim \mathcal{N}(x_0, 0)$; $P_{Y|X} \overset{?}{=} \delta(y - x_0)$

it's a delta (impulse) fn.

★ Relations btw joint, cond, marginal pdfs:

① $p_{y|x}(y|x) = \frac{p_{x,y}(x,y)}{p_x(x)} = \frac{p_{x,y}(x,y)}{\int_{y'} p_{x,y}(x,y')\,dy'}$

② $p_{x,y}(\,) = p_{x|y} \cdot p_y$
$= p_{y|x} \cdot p_x$

③ $p_y(y) = \int_{-\infty}^{\infty} \underbrace{p_{y|x}(y|x)\,p_x(x)}_{p_{x,y}}\,dx$

④ Bayes' Rule:

$$\underbrace{p_{y|x}(y|x)}_{\text{posterior pdf}} = \frac{\overbrace{p_{x|y}(x|y)}^{\text{likelihood}}\,\overbrace{p_y(y)}^{\text{prior prob.}}}{\underbrace{\int_{-\infty}^{\infty} p_{x|y}(x|y')\,p_y(y')\,dy'}_{p_x(x)\,=\,\text{evidence}}}$$

eg. Bayesian classifier.

8.8 (Reading) Modeling Human Learning

★ X, Y coin tossing experiment
$\rightarrow p(\text{Heads}) = p$.

\# heads in N trial

→ Bayesian analysis : $p_{y|x}(y|x) =$ Bayes rule.
calculate posterior pdf.

Conditional CDF:

- $F_{Y|X}(y|x) = P(Y \leq y \mid X = x)$

$$= \int_{-\infty}^{y} p_{Y|X}(y'|x)\, dy'$$

- If X & Y are independent:

$$p_{Y|X}(y|x) = p_Y(y)$$

$$F_{Y|X}(y|x) = F_Y(y)$$

e.g. previous example $\quad Y \mid X = x \sim N(\rho x, 1-\rho^2)$

$$F_{Y|X}(y|x) = 1 - \underbrace{Q\left(\frac{y - \rho x}{\sqrt{1-\rho^2}}\right)}_{\text{tail prob.}}$$

<u>Ex:</u> (13.2) 2 bulbs $\rightarrow$ X = lifetime of 1st bulb.

$\rightarrow$ Y = " " 2nd bulb (spare bulb).

Recall: time to failure is modeled by exponential pdf

$$X \sim \exp(\lambda).$$

e.g. expected time to failure of the bulb $= \frac{1}{\lambda^x}$

" " bulb 1 $\approx$ 100 hours

$\Rightarrow \lambda = 0.01$ /hours

X: in hours

$$p_X(x) = \lambda e^{-\lambda x} u(x) \quad ;$$

Y's $\rightarrow \lambda \Rightarrow$ expected time to failure 2nd bulb $\Rightarrow \frac{1}{\lambda^{y|x}} = \frac{1}{\lambda^x(1+\alpha x)}$

Given: e.g. $x = 30$ hours $\rightarrow \dfrac{1}{\lambda(1+\alpha\cdot 30)} \simeq 70$ hours

0.01 $\nearrow$ λ ; α: given 0.01 $\rightarrow$ parameter that puts in the wear-off (degradation) due to storage time.

$$p_{Y|X}(y|x) = \underbrace{(\lambda(1+\alpha x))}_{\lambda^{Y|X}} e^{-\lambda(1+\alpha x)y} u(y)$$

Goal:
Determine $\underbrace{p_Y(y)}_{\text{marginal pdf of } Y} = ?$

marginal pdf of Y (time to failure of the 2nd bulb spare.)

Start with $p_X(x)$
$\Rightarrow p_{Y|X}$
$\rightarrow p_Y(y) = ?$

$$p_Y(y) = \int_{-\infty}^{\infty} p_{Y|X}(y|x)\, p_X(x)\, dx$$

$$= \int_{\cancel{-\infty}\ 0}^{\infty} \lambda(1+\alpha x) e^{-\lambda(1+\alpha x)y} \cdot \lambda e^{-\lambda x} dx$$

$\vdots$ go through the derivation in the book

$$p_Y(y) = \lambda^2 \exp(-\lambda y)\left(\frac{1}{\lambda(\alpha y+1)} + \frac{\alpha}{[\lambda(\alpha y+1)]^2}\right) u(y)$$

Fig. 13.5