Proof: If part Let $\mu = E(X)$.

Now $(X-\mu)^2 = X^2 - 2\mu X + \mu^2$

$\Rightarrow$ By the thm (on linearity of expectation) stated in Pg 151, $(X-\mu)^2$ has finite mean and hence X has finite variance.

Also in this case,

$$\begin{aligned}
Var(X) &= E\left[(X-\mu)^2\right] \\
&= E\left[X^2 - 2\mu X + \mu^2\right] \\
&= E(X^2) - 2\mu E(X) + \mu^2 \\
&= E(X^2) - 2\mu^2 + \mu^2 \\
&= E(X^2) - \mu^2 \\
&= E(X^2) - (E(X))^2 .
\end{aligned}$$

Only if part

$$X^2 = (X-\mu)^2 + 2\mu X - \mu^2$$

$\Rightarrow$ By the thm stated in Pg 151, X has finite 2^{nd} moment.

Cor: If X has finite 2^{nd} moment, then X has finite variance, which is given by

$$\text{Var}(X) = E(X^2) - (E(X))^2.$$

Proof: Follows from Thm stated in ~~Pg 173~~ Pg 172 + the Exc stated at the end of Pg 162.

Cor: For any r.v. X, with finite 2^{nd} moment, $E(X^2) \geq (E(X))^2$.

Proof: $E(X^2) - (E(X))^2 = \text{Var}(X) = E[(X-\mu)^2] \geq 0$

$$\Rightarrow E(X^2) \geq (E(X))^2$$

Remark: Equality holds in the last corollary of Pg 174, i.e., $E(X^2) = (E(X))^2 < \infty$ if and only if $E[(X-\mu)^2] = 0$, which can be shown to be equivalent to $P(X=\mu)=1$, (i.e., ~~when~~ if and only if X is degenerate (constant)).

~~Exc~~

Exc: ① Suppose X is a discrete r.v. with finite 2^{nd} moment. Then show that

$E(X^2) = (E(X))^2$ holds if and only if X is a degenerate r.v.

② Suppose X is a cont r.v. with a cont pdf and ~~with~~ finite 2^{nd} moment. Then show that

$$E(X^2) > (E(X))^2.$$

Remark: The above exc essentially shows that for a discrete or cont r.v. X with finite 2^{nd} moment, $Var(X) \geq 0$ and $Var(X) = 0$ iff X is a degenerate. ~~r.v.~~ This result holds for any r.v. X.

Example: $X \sim \text{Gamma}(\alpha, \lambda)$. Compute $Var(X)$.

Solution: A pdf of X is

$$f_X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda x} x^{\alpha-1} \quad \text{if } x > 0.$$

Since X is a positive r.v., we have

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

$$= \int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda x} x^\alpha \, dx$$

$$= \frac{\Gamma(\alpha+1)}{\lambda \Gamma(\alpha)} \int_0^\infty \underbrace{\frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)} e^{-\lambda x} x^{(\alpha+1)-1}}_{\text{pdf of Gamma}(\alpha+1,\lambda)\text{ dist}^{n}} dx$$

$$= \frac{\alpha \Gamma(\alpha)}{\lambda \Gamma(\alpha)} = \frac{\alpha}{\lambda}.$$

Again,

$$E(X^2) = \int_0^\infty x^2 f_X(x)\, dx$$

$$= \int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda x} x^{\alpha+1}\, dx$$

$$= \frac{\Gamma(\alpha+2)}{\lambda^2 \Gamma(\alpha)} \int_0^\infty \underbrace{\frac{\lambda^{\alpha+2}}{\Gamma(\alpha+2)} e^{-\lambda x} x^{(\alpha+2)-1}}_{\text{pdf of Gamma}(\alpha+2,\lambda)\text{ distn.}} dx$$

$$= \frac{\Gamma(\alpha+2)}{\lambda^2 \Gamma(\alpha)}$$

$$= \frac{\alpha(\alpha+1)}{\lambda^2} < \infty. \quad [\because \Gamma(\alpha+2) = \alpha(\alpha+1)\Gamma(\alpha)]$$

Therefore, X has finite variance by the first corollary of Pg 174 and

$$Var(X) = E(X^2) - (E(X))^2$$

$$= \frac{\alpha^2+\alpha}{\lambda^2} - \frac{\alpha^2}{\lambda^2} = \frac{\alpha}{\lambda^2}.$$

$\therefore$ We have thus shown the following.

<u>Fact</u>: $X \sim \text{Gamma}(\alpha, \lambda) \Rightarrow E(X) = \frac{\alpha}{\lambda}$, $V(X) = \frac{\alpha}{\lambda^2}$.

<u>Remark</u>: When $n = \alpha \in \mathbb{N}$, we can compute $E(X)$ using linearity of expectation as follows.

Suppose $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Exp}(\lambda)$.

Then $X = \sum_{i=1}^{n} X_i \sim \text{Gamma}(n, \lambda)$

$$\Rightarrow E(X) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} \frac{1}{\lambda} = \frac{n}{\lambda}.$$

Again, the above calculation did not use the independence of X_i's — we shall use it when $Var(X)$ is computed using the addivity of variance under independence.

<u>Exc</u>: Suppose $X \sim \text{Beta}(a, b)$. Compute $E(X)$ and $Var(X)$. (Simplification is not needed)

[<u>Hint</u>: You may use $B(r, s) = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}$.]

Covariance and Correlation as Measures of Association

Suppose (X, Y) is a random vector and we would like to measure the amount of association between X and Y.

The most basic such measure is called covariance, which is defined below.

Defn: Suppose X and Y are jointly ~~distributed~~ distributed r.v.s with finite means μ_X and μ_Y, respectively. Then we say that X and Y have have finite covariance if $(X-\mu_X)(Y-\mu_Y)$ has finite mean. And in this case, we define

Covariance of X and Y

$$= \text{Cov}(X, Y) = E\left[(X-\mu_X)(Y-\mu_Y)\right].$$

Remarks: ① If $X \equiv Y$, that is if $P(X=Y)=1$, then

$$\text{Cov}(X, Y) = \text{Cov}(X, X) = \text{Var}(X).$$

2 Interpretation of Covariance

a If X and Y are positively associated, i.e., if a lower value of X tend to give rise to a lower value of Y and a higher value of X tend to give rise to a higher value of Y, then $Cov(X,Y)$ is going to be positive. This is because whenever $X > \mu_X$, we will have $Y > \mu_Y$ ~~mos~~ with higher probability and whenever $X < \mu_X$, we will have $Y < \mu_Y$ with higher probability, and therefore the product $(X-\mu_X)(Y-\mu_Y)$ will be positive ~~most~~ with high probability making $Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$ positive as well. This will be the case in the example of drainage network model given in Pg 15 – 18; see also the ~~re~~ remark in Pg 19.

(b) If X and Y are negatively associated, i.e., if lower value of one of them tend to give rise to a higher value of the other and vice-versa, then $Cov(X,Y)$ is going to be negative. The reason is ~~very si~~ analogous to the one given in Pg 180 except that the product $(X-\mu_X)(Y-\mu_Y)$ will be more likely to be negative leading to its expectation, i.e., $Cov(X,Y)$, ~~being~~ to be negative as well. This will be the case in the example on Polya's urn scheme given in Pg 9 - 13; see also ~~Re~~ Note 3 of Pg 14.

(c) Note that (a) and (b) above give the ~~inter~~ interpretation of the sign of covariance. The ~~interpretation~~ interpretation of the value of covariance is tricky. While covariance is a measure of association, it does get affected by scaling (For example, $Cov(27X, 42Y) = (27 \times 42)\, Cov(X,Y)$) making it difficult to understand and interpret the ~~meani~~ value of covariance, which is clearly not unit-free.