AALBORG UNIVERSITET

Exam notes

4th semester: *Probability theory*

Kasper Rosenkrands

Contents

1 Basics of probability (incl. combinatorics, law of total probability and Bayes' formula). 1

2 Discrete stochastic variables and distributions (incl. means and variances). 2

3 Continuous stochastic variables and distributions (incl. means and variances). 4

4 Two random variables: select from topics such as joint distribution, conditional distribution, independence and convolution. 7

4.1 Convolution 8

5 Two random variables: select from topics such as covariance and correlation, conditional expectation, conditional variance and the bivariate normal distribution. 9

5.1 Conditional Expectation as a Random Variable 9

5.2 Conditional Variance 10

5.3 Covarince and Correlation 10

5.4 The Correlation Coefficient 11

6 Generating functions (possibly with a focus on how probability generating functions relate to thinning of a Poisson process). 13

6.1 The Probability Generating Function 13

6.2 The Moment Generating Function 14

6.3 The Poisson Proces 15

7 Limit theorems. 17

8 Markov chains. 18

9 Stochastic simulation. 19

1 Basics of probability (incl. combinatorics, law of total probability and Bayes' formula).

Theorem 1.1 (Law of Total Probability). Let $B_1, B_2, \ldots$ be a sequence of events such that

(a) $P(B_k) > 0$ for $k = 1, 2, \ldots$

(b) B_i og B_j are disjoint whenever $i \neq j$

(c) $S = \bigcup_{k=1}^{\infty} B_k$

Then, for any event A, we have

$$P(A) = \sum_{k=1}^{\infty} P(A|B_k)P(B_k).$$

Proof. First note that

$$A = A \cap S = \bigcup_{k=1}^{\infty} (A \cap B_k),$$

by the distributive law for infinite unions. Since $A \cap B_1, A \cap B_2, \ldots$ are pairwise disjoint, we get

$$P(A) = P(A \cap B_k) = \sum_{k=1}^{\infty} P(A|B_k)P(B_k).$$

Which proves the theorem. Note that the result also holds for finite sequences. ■

Corollary 1.6. If $0 < P(B) < 1$, for $B \subseteq S$, then

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$$

Proposition 1.11 (Bayes' Formula). Under the same assumptions as in the law of total probability and if $P(A) > 0$, then for any event B_j, we have

$$\begin{aligned} P(B_j|A) &= \frac{P(A|B_j)P(B_j)}{\sum_{k=1}^{\infty} P(A|B_k)P(B_k)} \\ &= \frac{P(A|B_j)P(B_j)}{P(A)} \qquad \text{according to the LTP.} \end{aligned}$$

2 Discrete stochastic variables and distributions (incl. means and variances).

Definition 2.1. A random variable is a real random variable that gets its values from a random experiment.

$$X: \; S \to \mathbb{R}.$$

Definition 2.2. If the range of X is countable, then X is called a *discrete random variable.*

Definition 2.3. Let X be a discrete random variable with range $\{x_1, x_2, \ldots\}$ (finite or countably infinite). The function

$$p(x_k) = P(X = x_k), \quad k = 1, 2, \ldots$$

is called the *probability mass function* (pmf) of X.

Proposition 2.1. A function p is a possible pmf of a discrete random variable on the range $\{1, 2, \ldots\}$ if and only if

(a) $p(x_k) \geq 0 \quad$ for $k = 1, 2, \ldots$

(b) $\sum_{k=1}^{\infty} p(x_k) = 1$

Definition 2.4. Let X be any random variable. The function

$$F(x) = P(X \leq x), \quad \text{for } x \in \mathbb{R},$$

is called the (*cumulative*) *distribution function* (cdf) of X.

Defintion 2.8. Let X be a discrete random variable with range $\{x_1, x_2, \ldots\}$ (finite or countably infinite) and probability mass function p. The *expected value* of X is defined as

$$E[X] = \sum_{k=1}^{\infty} x_k p(x_k).$$

Proposition 2.9. Let X be a discrete random variable with range $\{0, 1, \ldots\}$. Then

$$E[X] = \sum_{n=0}^{\infty} P(X > n)$$

Proof. Note that $k = \sum_{n=1}^{k} 1$, and use the definition of expected value to obtain

$$\begin{aligned} E[X] &= \sum_{k=1}^{\infty} kP(X = k) = \sum_{k=1}^{\infty} \sum_{n=1}^{k} P(X = k) \\ &= \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} P(X = k) = \sum_{n=1}^{\infty} P(X \geq n) \\ &= \sum_{n=0}^{\infty} P(X > n) \end{aligned}$$

■

Proposition 2.11 (Linearity for the Expectation). Let X be any random variable, and let a and b be real numbers. Then

$$E[aX + b] = aE[X] + b$$

Proof. We prove this in the discrete case, for $a > 0$. Let $Y = aX + b$, and note that Y is a discrete random variable and by definition 2.8 expected value

$$\begin{aligned} E[Y] &= \sum_{k=1}^{\infty} y_k p_Y(y_k) = \sum_{k=1}^{\infty} y_k p_X\left(\frac{y_k - b}{a}\right) \\ &= \sum_{k=1}^{\infty} (ax_k + b) p_X(x_k) \\ &= a \sum_{k=1}^{\infty} x_k p_X(x_k) + b \sum_{k=1}^{\infty} p_X(x_k) \\ &= aE[X] + b \end{aligned}$$

and we are done. **Måske skal man bruge 2.12 her (side 102).** ■

3 Continuous stochastic variables and distributions (incl. means and variances).

Definition 2.5. If the cdf F is a continuous and differentiable function, then X is said to be a *continuous random variable.*

Proposition 2.3. If F is the cdf of any random variable, F has the following properties:

(a) It is nondecreasing

(b) It is right-continuous

(c) It has the limits $F(-\infty) = 0$ and $F(\infty) = 1$ (where the limits may or may not be attained at finite x).

Proposition 2.4. Let X be any random variable with cdf F. Then

(a) $P(a < X \leq b) = F(b) - F(a), \quad a \leq b$

(b) $P(X > x) = 1 - F(x), \quad x \in \mathbb{R}$

(c) F is continuous from the right with limit form the left and increasing

Definition 2.6. The function $f(x) = F'(x)$ is called the *probability density function* (pdf) of X.

Proposition 2.5. Let X be a continuous random variable with pdf f and cdf F. Then

(a) $F(x) = \int_{-\infty}^{x} f(t)dt, \quad x \in \mathbb{R}$

(b) $f(x) = F'(x), \quad x \in \mathbb{R}$

(c) For $B \subseteq \mathbb{R}, \quad P(X \in B) = \int_B f(x)dx$

Proposition 2.6. A function f is a possible pdf of some continuos random variable if and only if

(a) $f(x) \geq 0, \quad x \in \mathbb{R}$

(b) $\int_{-\infty}^{\infty} f(x)dx = 1$

Proposition 2.8. Let X be a continuous random variable with pdf X, let g be a strictly increasing or strictly decreasing, differentiable function, and let $Y = g(X)$. Then Y has pdf

$$f_Y(y) = \left| \frac{d}{dy} g^{-1}(y) \right| f_X(g^{-1}(y))$$

for y in range of Y.

Defintion 2.9. Let X be a continuous random variable with pdf f. The *expected value* of X is defined as

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx = \int_{\mathbb{R}} xf(x)dx,$$

notice that the last equality is not always satisfied, but for the purpose of this course it is.

Proposition 2.10. Let X be a continuous random variable with range $[0, \infty)$. Then

$$E[x] = \int_{0}^{\infty} P(X > x)dx$$

Proposition 2.11 (Linearity for the Expectation). Let X be any random variable, and let a and b be real numbers. Then

$$E[aX + b] = aE[X] + b$$

Proof. We prove this in the continuous case, for $a > 0$. Let $Y = aX + b$, and note that Y is a continuous random variable, which by proposition 2.8 has pdf

$$f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right)$$

and by definition, the expected value of Y is

$$E[Y] = \int_{-\infty}^{\infty} yf_Y(y)dy = \frac{1}{a}\int_{-\infty}^{\infty} yf_X\left(\frac{y-b}{a}\right)dy$$

where the variable $y = ax + b$ gives $dy = a\ dx$ and hence

$$\begin{aligned} E[Y] &= \int_{-\infty}^{\infty} (ax+b)f_X(x)dx \\ &= a\int_{-\infty}^{\infty} xf_X(x)dx + b\int_{-\infty}^{\infty} f_X(x)dx \\ &= aE[X] + b \end{aligned}$$

and we are done. ■

Definition 2.10. Let X be a random variable with expected value μ. The *variance* of X is defined as

$$Var[X] = E\left[(X - \mu)^2\right]$$

Definition 2.11. Let X be a random variable with variance $\sigma^2 = Var[X]$. The *standard deviation* of X is then defined as $\sigma = \sqrt{Var[X]}$.

Corollary 2.2.

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

Proof. By proposition 2.12, we have

$$\begin{aligned}
Var[X] &= E\Big[(X-\mu)^2\Big] = \int_{-\infty}^{\infty} (x-\mu)^2 dx \\
&= \int_{-\infty}^{\infty} (x^2 - 2x\mu + \mu^2) f(x) dx \\
&= \int_{-\infty}^{\infty} x^2 f(x) dx - 2\mu \int_{-\infty}^{\infty} x f(x) dx + \mu^2 \int_{-\infty}^{\infty} f(x) dx \\
&= E[X^2] - 2\mu E[X] + \mu^2 \\
&= E[X^2] - 2E[X]^2 + E[X]^2 \\
&= E[X^2] - E[X]^2.
\end{aligned}$$

■

Proposition 2.14 (Chebyshev's Inequality). Let X be any random variable with mean μ and variance σ^2. For any consatn $c > 0$, we have

$$P(|X-\mu| \geq c\sigma) \leq \frac{1}{c^2}.$$

Proof. Let us prove the continuous case. Fix c and let B be the set $\{x \in \mathbb{R} : |x-\mu| \geq c\sigma\}$. We get

$$\begin{aligned}
\sigma^2 = E[(X-\mu)^2] &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx = \int_{\mathbb{R}} (x-\mu)^2 f(x) dx \\
&\geq \int_B (x-\mu)^2 f(x) dx \geq c^2\sigma^2 \int_B f(x) dx = c^2\sigma^2 P(X \in B).
\end{aligned}$$

Which gives the desired inequality. ■

4 Two random variables: select from topics such as joint distribution, conditional distribution, independence and convolution.

Definition 3.1. Let X and Y be random variables. The pair (X, Y) is then called a (two-dimensional) *random vector.*

Definition 3.2. The *joint distribution function* (joint cdf) of (X, Y) is defined as

$$F(x, y) = P(X \leq x, Y \leq y)$$

for $x, y \in \mathbb{R}$.

Proposition 3.1 (Marginal cdf). If (X, Y) has joint cdf F, then X and Y have cdfs

$$F_X(x) = F(x, \infty) \quad \text{and} \quad F_Y(y) = F(\infty, y)$$

for $x, y \in \mathbb{R}$. Notice that there is a slight abuse of notation here, $F(x, \infty)$ refers to $\lim_{y \to \infty} F(x, y)$.

Definition 3.5. If there exists a function f such that

$$P((X, Y) \in B) = \int \int_B f(x, y) dx dy$$

for all subsets $B \subseteq \mathbb{R}^2$, then X adn Y are said to be *jointly continuous.* The function f is calle the *joint pdf.*

Propostion 3.3. If X and Y are jointly continuous with joint cdf F adn joint pdf f, then

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y), \quad x, y \in \mathbb{R}.$$

Proposition 3.4. A function f is a possible joint pdf for the random variables X and Y if and only if

(a) $f(x, y) \geq 0$ for all $x, y \in \mathbb{R}$

(b) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx\ dy = 1$

Proposition 3.5. Suppose that X and Y are jointly continuous with joint pdf f. Then X adn Y are continuous random variables with marginal pdfs

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy, \quad x \in \mathbb{R}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx, \quad y \in \mathbb{R}.$$

Definition 3.7. Let (X, Y) ne jointly continuous with joint pdf f. The *conditional pdf of Y given $X = x$* is defined as

$$f_Y(y|x) = \frac{f(x,y)}{f_X(x)}, \quad y \in \mathbb{R}.$$

The following proposition is a continuous version of the law of total probability.

Proposition 3.6. Let X and Y be jointly continuous. Then

(a)

$$f_Y(y) = \int_{-\infty}^{\infty} f_Y(y|x) f_X(x) dx, \quad y \in \mathbb{R}$$

(b)

$$P(Y \in B) = \int_{-\infty}^{\infty} P(Y \in B|X = x) f_X(x) dx, \quad B \subseteq \mathbb{R}.$$

Proof. For (a), just combine Proposition 3.5 with the definition of conditional pdf for (b), part (a) gives

$$\begin{aligned} P(Y \in B) &= \int_B f_Y(y) dy = \int_B \int_{-\infty}^{\infty} f_Y(y|x) f_X(x) dx \; dy \\ &= \int_{-\infty}^{\infty} \int_B f_Y(y|x) f_X(x) dy \; dx \\ &= \int_{-\infty}^{\infty} P(Y \in B|X = x) f_X(x) dx \end{aligned}$$

as desired. ∎

Proposition 3.10. Suppose that X and Y are jointly continuous with joint pdf f. Then X and Y are independent if and only if

$$f(x, y) = f_X(x) f_Y(y)$$

for all $x, y \in \mathbb{R}$.

Corrollary 3.1. *The random variables X and Y are independent if and only if "the joint is the product of the marginals."*

4.1 Convolution

Proposition 3.34. Let X and Y be independent continuous random variables with pdfs f_X and f_Y, respectively. The pdf of the sum $X + Y$ is then

$$f_{X+Y} = \int_{-\infty}^{\infty} f_Y(x - u) f_X(u) du, \quad x \in \mathbb{R}.$$

5 Two random variables: select from topics such as covariance and correlation, conditional expectation, conditional variance and the bivariate normal distribution.

Definition 3.10. Let y be random variable and B an event with $P(B) > 0$. The *conditional expectation* of Y given B is defined as

$$E[Y|B] = \begin{cases} \sum_{k=1}^{\infty} y_k P(Y = y_k|B) & \text{if } Y \text{ is discrete with range } \{y_1, y_2, \ldots\} \\ \int_{-\infty}^{\infty} y f_Y(y|B) dy & \text{if } Y \text{ is continuous} \end{cases}$$

Definition 3.11. Suppose that X and Y are discrete. We define

$$E[Y|X = x_j] = \sum_{k=1}^{\infty} y_k p_Y(y_k|x_j).$$

Definition 3.12. Suppose that X and Y are jointly continuous. We define

$$E[Y|X = x] = \int_{-\infty}^{\infty} y f_Y(y|x) dx.$$

Following the usual intuitive interpretation, this is the expected value for Y if we know that $X = x$. The law of total expectation now takes the following form.

S.uppose that X and Y are jointly continuous. Then

$$E[Y] = \int_{-\infty}^{\infty} E[Y|X = x] f_X(x) dx.$$

5.1 Conditional Expectation as a Random Variable

Definition 3.13. The *conditional expectation* of Y given X, $E[Y|X]$, is a random variable that equals $E[Y|X = x]$ whenever $X = x$.

Corollary 3.5.

$$E[Y] = E\Big[E[Y|X]\Big]$$

Definition 3.14. Let $g(X)$ be a predictor of Y. The *mean square error* is defined as

$$E\Big[(Y - g(X))^2\Big].$$

Proposition 3.18. Among all predictors $g(X)$ of Y, the mean square error is minimized by $E[Y|X]$.

We omit the proof and instead refer to an intuitive argument. Suppose that we want to predcit Y as much as possible by a constant value c. Then, we want to minimize $E[(Y-c)^2]$, and with $\mu = E[Y]$ we get

$$\begin{aligned} E\Big[(Y-c)^2\Big] &= E\Big[(Y-\mu+\mu-c)^2\Big] \\ &= E\Big[(Y-\mu)^2] + 2(\mu-c)E[(Y-\mu)] + (\mu-c)^2 \\ &= Var[Y] + (\mu-c)^2 \end{aligned}$$

since $E[Y-\mu]=0$. But the last expression is minimized when $\mu = c$ and hence μ is the best predictor of Y among all constants. This is not too surprising; if we do not know anything about Y, the best guess should be the expected value $E[Y]$. Now, if we observe another random variable X, the same ought to be true: Y is best predicted by its expected value given the random variable X, that is, $E[Y|X]$.

5.2 Conditional Variance

Definition 3.15. The *conditional variance* of Y given X is defined as

$$Var[Y|X] = E\Big[(Y - E[Y|X])^2|X\Big].$$

Note that the conditonal varince is also a random variable and we think of it as the variance of Y given the value X. In particular, if we have the observed $X = x$, then we can denote and defined

$$Var[Y|X=x] = E\Big[(Y - E[Y|X=x])^2|X=x\Big].$$

also note that if X and Y are independent, $E[Y|X] = E[Y]$, and the definition boils down to the regular variance. There is an analog of Corollary 2.2, which we leave to the reader to prove.

Corollary 3.7.

$$Var[Y|X] = E\Big[[Y^2|X] - (E[Y|X])^2$$

There is also a "law of total variance", which looks slightly more complicated than that of total expectation.

Proposition 3.19.

$$Var[Y] = Var\Big[E[Y|X]\Big] + E\Big[Var[Y|X]\Big]$$

5.3 Covarince and Correlation

Definition 3.16. The *covariance* of X and Y is defined as

$$Cov[X,Y] = E\Big[(X - E[X])(Y - E[Y])\Big].$$

Proposition 3.20.

$$Cov[X,Y] = E[XY] - E[X]E[Y]$$

Corollary 3.8. *If X and Y are independent, then $Cov[X,Y] = 0$.*

Proposition 3.21.

$$Var[X+Y = Var[x] + Var[Y] + 2Cov[X,Y].$$

Proof. By the definition of variance and covariance and repeated use of properties of expected values, we get

$$\begin{aligned} Var[X+Y] &= E\Big[(X+Y-E[X+Y])^2] \\ &= E\Big[(X-E[X]+Y-E[Y])^2] \\ &= E\Big[(X-E[X])^2 + (Y-E[Y])^2 + 2(X-E[X])(Y-E[Y])] \\ &= Var[X] + Var[Y] + Cov[X,Y] \end{aligned}$$

and we are done. ∎

Proposition 3.22. Let X, Y, and Z be random variables and let a and b be real numbers. Then

(a) $Cov[X,X] = Var[X]$

(b) $Cov[aX,bX] = abCov[X,Y]$

(c) $Cov[X+Y,Z] = Cov[X,Z] + Cov[Y,Z]$

together these properties indicate that covariance is *bilinear*.

5.4 The Correlation Coefficient

Definition 3.17. The *correlation coefficient* of X and Y is defined as

$$\rho(X,Y) = \frac{Cov[X,Y]}{\sqrt{Var[X]Var[Y]}}.$$

The correlation coefficient is dimensionless. To demonstrate this, take $a, b > 0$ and note that

$$\begin{aligned} \rho(aX,bY) &= \frac{Cov[aX,bY]}{\sqrt{Var[aX]Var[bY]}} \\ &= \frac{abCov[X,Y]}{\sqrt{a^2Var[X]b^2Var[Y]}} = \rho(X,Y). \end{aligned}$$

We also call $\rho(X,Y)$ simply the correlation between X and Y. Here are som good properties of the correlation coefficient.

Proposition 3.25. The correlation coefficient of any pair of random variables X and Y satisfies

(a) $-1 \leq \rho(X, Y) \leq 1$

(b) If X and Y are independent, then $\rho(X, Y) = 1$

(c) $\rho(X, Y) = 1$ if and only if $Y = aX + b$, where $a > 0$

(d) $\rho(X, Y) = -1$ if and only if $Y = aX + b$, where $a < 0$

Proof. Let $Var[X] = \sigma_1^2$ and $Var[Y] = \sigma_2^2$. For (a), first apply Proposition 3.21 to the random variables X/σ_1 and Y/σ_2 and use the properties of the variance and covariance to obtain

$$0 \leq Var\left[\frac{X}{\sigma_1} + \frac{Y}{\sigma_2}\right] = \frac{Var[X]}{\sigma_1} + \frac{Var[Y]}{\sigma_2} + \frac{2Cov[X,Y]}{\sigma_1 \sigma_2} = 2 + 2\rho$$

which gives $\rho \geq 1$. To show that $\rho \leq 1$, instead use X/σ_1 and $-Y/\sigma_2$. Part (b) follows from Corollary 3.8, and parts (c) and (d) follows from Proposition 2.16, applied to tha random variables $X/\sigma_1 - Y/\sigma_2$ and $X/\sigma_1 + Y/\sigma_2$, respectively. Note that this also gives a and b expressed in terms of the means, variances, and correlation coefficeinet (see problem 90). ■

Her mangler der bivariate normal!

6 Generating functions (possibly with a focus on how probability generating functions relate to thinning of a Poisson process).

Generating functions, or *transforms*, are very useful in probability theory as in other fields of mathematics. Several different generating functions are used, depending on the type of random variable. We will discuss two, one that is useful for discrete random variables and the other for continuous random variables.

6.1 The Probability Generating Function

When we study nonnegative, integer-valued random variables, the following function proves to be a useful tool.

Definition 3.23. Let X be nonnegative and integer valued. The function

$$G_X(s) = E[s^X] = \sum_{k=0}^{\infty} s^k P(X = k) \quad 0 \leq s \leq 1$$

is called the *probability generating function* (pgf) of X.

Corollary 3.12. Let X be a nonnegative and integer valued with pgf G_X. then

$$G_X(0) = p_X(0) \quad \text{and} \quad G_X(1) = 1$$

Proposition 3.36. Let X be nonnegative and integer valued with pgf G_X. then

$$p_X(k) = \frac{G_X^{(k)}(0)}{k!}, \quad k = 0, 1, \ldots$$

where $G_X^{(k)}(0)$ denotes the kth derivative of G_X.

Proposition 3.37. If X has pgf G_X, Then

$$E[X] = G'_X(1) \quad \text{and} \quad Var[X] = G''_X(1) + G'_X(1) - G'_X(1)^2$$

Proposition 3.38. Let $X_1, X_2, \ldots, X_n$ be independent random variables with pgfs $G_1, G_2, \ldots, G_n$ respectively and let $S_n = X_1 + X_2 + \cdots + X_n$. Then S_n has pgf

$$G_{S_n}(s) = G_1(s)G_2(s)\cdots G_n(s), \quad 0 \leq s \leq 1$$

Proof. Since $X_1, \ldots, X_n$ are independent, the random variables $s^{X_1}, \ldots, s^{X_n}$ are also independent for each s in [0,1], and we get

$$\begin{aligned} G_{S_n}(s) &= E[s^{X_1+X_2+\cdots+X_n}] \\ &= E[s^{X_1}]E[s^{X_2}]\cdots E[s^{X_n}] \\ &= G_1(s)G_2(s)\cdots G_n(s) \end{aligned}$$

and we are done. ∎

Proposition 3.39. Let $X_1, X_2, \ldots$ be i.i.d. nonnegative and integer valued with commen pgf G_X, and let N be nonnegative and integer valued, and independent of the X_k, with pgf G_N. Then $S_N = X_1 + \cdots + X_N$ has pgf

$$G_{S_N}(s) = G_N(G_X(s))$$

the composition of G_N and G_X.

Proof. We condition on N to obtain

$$\begin{aligned} G_{S_N}(s) &= E\left[E\left[s^{S_N}|N=n\right]\right] \\ &= \sum_{n=0}^{\infty} E[s^{S_N}|N=n]P(N=n) \\ &= \sum_{n=0}^{\infty} E[s^{S_N}]P(N=n) \end{aligned}$$

since N and S_n are independent. Now note that

$$E[s^{S_N}] = G_X(s)^n$$

by Proposition 3.38 and we get

$$G_{S_N}(s) = \sum_{n=0}^{\infty} G_X(s)^n P(N=n) = G_N(G_X(s))$$

the pgf of N evaluated at the point $G_X(s)$. ■

Corollary 3.13. Under the assumptions of Proposition 3.39, it holds that

$$\begin{aligned} E[S_N] &= E[N]\mu \\ Var[S_N] &= E[N]\sigma^2 + Var[N]\mu^2 \end{aligned}$$

where $\mu = E[X_k]$ and $\sigma^2 = Var[X_k]$.

6.2 The Moment Generating Function

The probability generating function is an excellent tool for nonnegative and integer-valued random variables. For other random variables, we can instead use the following more general generating function.

Definition 3.24. Let X be a random variable. The function

$$M_X(t) = E[e^{tX}], \quad t \in \mathbb{R}$$

is calle the *moment generating function* (mgf) of X.

If X is continuous with pdf f_X, we compute the mgf by

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)dx$$

and for discrete X, we get a sum instead. Note that if X is nonnegative integer valued with pgf G_X, then

$$M_X(t) = G_X(e^t)$$

which immediately give the mgf for the distributions for which we computed the pgf above.

Corollary 3.15. If X has mgf $M_X(t)$, then

$$E[X] = M'_X(0) \quad \text{and} \quad Var[X] = M''_X(0) - (M'_X(0))^2$$

Proof. By differentiating $M_X(t)$ with respect to t, we get

$$M'_X(t) = \frac{d}{dt}E[e^{tx}] = E[Xe^{tx}]$$

where we assumed that we can interchange differentiation and expectation. Note how t is the variable of differentiation and we view X as fixed. In the case of a discrete X, this amounts to differentiating a sum termwise, and in the case of a continuous X, it means that we can differentiate under the integral sign. This is by no means obvious but can be verified. We will not address this issue further. Differentiating once more gives

$$M''_X(t) = \frac{d}{dt}E[Xe^{tx}] = E[X^2e^{tx}]$$

which gives

$$Var[X] = E[X^2] - (E[X])^2 = M''_X(0) - (M'_X(0))^2$$

as desired. ■

Note we get the general result

$$E[X^n] = M_X^{(n)}(0), \quad n = 1, 2, \ldots$$

where $M_X^{(n)}$ is the nth derivative of M_X. The number $E[X^n]$ is called the nth *moment* of X, hence the term moment generating function. Compare it with the probability generating function that generates probabilities by differentiating and setting $s = 0$. The moment generating function also turns sum into products, according to the following proposition, which you may prove as an excercise.

Proposition 3.40. Let $X_1, X_2, \ldots, X_n$ be independent random variables with mgfs $M_1, M_2, \ldots, M_n$, respectively, and let $S_n = X_1 + \cdots + X_n$. Then S_n has mgfs

$$M_{S_n}(t) = M_1(t)M_2(t)\cdots M_n(t), \quad t \in \mathbb{R}$$

6.3 The Poisson Proces

Definition 3.25. A point process where times between consecutive points are i.i.d. tandom variiables that are $\exp(\lambda)$ is called a *Poisson process* with rate λ.

Proposition 3.41. Consider a Poisson process with rate λ, where $X(t)$ is the number of points in an interval of length t. Then

$$X(t)\ \mathrm{Poi}(\lambda t)$$

Recall that the parameter in the Poisson distribution is also the expected value. Hence, we have

$$E[X(t)] = \lambda t$$

which makes sense since λ is the mean number of points per time unit and t is the length of the time interval. In practical applications, we need to be careful to use the same time units for λ and t.

Proposition 3.42. In a Poisson process with rate λ

(a) $T_1, T_2, \ldots$ are independent and $\exp(\lambda)$

(b) $X(T_1), X(T_2), \ldots$ are independent and $X(t_j)\ Poi(\lambda t_j), \quad j = 1, 2, \ldots$

7 Limit theorems.

8 Markov chains.

9 Stochastic simulation.