Formulary Advanced Statistics I - Winter term 2023/2024

1. Elements of probability theory

Theorem 1.1 *Let A be an event in S. Then $P(A) = 1 - P(\bar{A})$.*

Theorem 1.2 $P(\emptyset) = 0$.

Theorem 1.3 *Let A and B be events in S such that $A \subset B$. Then $P(A) \leq P(B)$ and $P(B - A) = P(B) - P(A)$.*

Theorem 1.4 *Let A and B be events in S. Then $P(A) = P(A \cap B) + P(A \cap \bar{B})$.*

Theorem 1.5 *Let A and B be events in S. Then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.*

Corollary 1.1 (Boole's Inequality) $P(A \cup B) \leq P(A) + P(B)$.

Theorem 1.6 *Let A be an event in S. Then $P(A) \in [0, 1]$.*

Theorem 1.7 (Bonferroni's Inequality) *Let A and B be events in S. Then $P(A \cap B) \geq 1 - P(\bar{A}) - P(\bar{B})$.*

Theorem 1.8 *Let $A_1, \ldots, A_n$ be events in S. Then $P(\cap_{i=1}^n A_i) \geq 1 - \sum_{i=1}^n P(\bar{A}_i)$ and $P(\cup_{i=1}^n A_i) \leq 1 - \sum_{i=1}^n P(A_i)$.*

Theorem 1.9 (Classical probability) *Let S be the finite sample space for an experiment having $n = N(S)$ equally likely outcomes, say $E_1, ..., E_n$, and let $A \subset S$ be an event containing $N(A)$ elements. Then the probability of the event A is given by $N(A)/N(S)$.*

Theorem 1.10 *Given a probability space $\{S, Y, P\}$ and an event B for which $P(B) \neq 0$, $P(A \mid B) = P(A \cap B)/P(B)$ defines a probability set function with domain Y.*

Theorem 1.11 (Multiplication Rule) *Let A and B be any two events in S for which $P(B) \neq 0$. Then $P(A \cap B) = P(A \mid B)P(B)$.*

Theorem 1.12 (Extended Multiplication Rule) *Let $A_1, A_2, \ldots, A_n, n \geq 2$, be events in S. Then if all of the conditional probabilities exist, $P(\cap_{i=1}^n A_i) = P(A_1) \cdot P(A_2|A_1) \cdot \ldots \cdot P(A_n|A_{n-1} \cap A_{n-2} \cap \ldots \cap A_1) = P(A_1) \prod_{i=2}^n P\left(A_i \mid \cap_{j=1}^{i-1} A_j\right)$.*

Theorem 1.13 *If events A and B are independent, then events A and $\bar{B}$, $\bar{A}$ and B, and $\bar{A}$ and $\bar{B}$ are also independent.*

Theorem 1.14 (Law of Total Probability) *Let the events $B_i, i \in I$, be a finite or countably infinite partition of S, so that $B_j \cap B_k = \emptyset$ for $j \neq k$, and $\cup_{i \in I} B_i = S$. Let $P(B_i) > 0 \; \forall \; i \in I$. Then total probability of event A is $P(A) = \sum_{i \in I} P(A \mid B_i) P(B_i)$.*

Corollary 1.2 (Bayes's Rule) *Let the events $B_i, i \in I$, be a finite or countably infinite partition of S, so that $B_j \cap B_k = \emptyset$ for $j \neq k$ and $\cup_{i \in I} B_i = S$. Let $P(B_i) > 0 \; \forall \; i \in I$. Then, provided $P(A) \neq 0$, $P(B_j \mid A) = \dfrac{P(A \mid B_j) P(B_j)}{\sum_{i \in I} P(A \mid B_i) P(B_i)} \; \forall j \in I$.*

2. Random variables and their probability distributions

Theorem 2.1 (Properties of a cdf) *For any cdf F, it holds that: (i) $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$; (ii) $F(x)$ is a non decreasing function on x; that is, $F(a) \leq F(b)$ for $a < b$; (iii) $F(x)$ is right-continuous; that is, $\lim_{h \downarrow 0} F(x + h) = F(x)$.*

Theorem 2.2 *Let $x_1 < x_2 < x_3 < \cdots$ be the countable set of outcomes in the range of the discrete random variable X. Then the pdf for X obtains as $f(x_i) = \begin{cases} F(x_i), & i = 1 \\ F(x_i) - F(x_{i-1}), & i = 2, 3, \ldots \end{cases}$ and 0 if $x \notin R(X)$.*

Theorem 2.3 *Let $f(x)$ and $F(x)$ denote the pdf and cdf of a continuous random variable X. Then the pdf for X obtains as $f(x) = \frac{dF(x)}{dx}$, whenever $f(x)$ is continuous, and 0 elsewhere.*

Theorem 2.4 (Properties of joint cdfs) *For any multivariate cdf F, it holds that: (i)* $\lim_{b_i \to -\infty} F(b_1, ..., b_n) = P_X(\emptyset) = 0, \quad$ *for any* $\quad i = 1, ..., n$*; (ii)* $\lim_{b_i \to \infty, \forall i} F(b_1, ..., b_n) = P_{\boldsymbol{X}}(R(\boldsymbol{X})) = 1$*; (iii) F is a non decreasing function on $(x_1, ..., x_n)$, that is, $F(a) \leq F(b)$ for (the suitably defined vector inequality) $a = (a_1 \dots a_n)' < (b_1 \dots b_n) = b$; (iv) Discrete joint cdfs have a countable number of jump discontinuities and joint cdfs for continuous random variables are continuous without jump discontinuities.*

Theorem 2.5 *Let (X, Y) be a discrete bivariate random variable with joint cdf $F(x, y)$ and range $R(X, Y) = \{x_1 < x_2 < x_3 < \cdots, y_1 < y_2 < y_3 < \cdots\}$. Then the joint pdf obtains as $f(x_1, y_1) = F(x_1, y_1)$, $f(x_1, y_j) = F(x_1, y_j) - F(x_1, y_{j-1}), \quad j \geq 2$, $f(x_i, y_1) = F(x_i, y_1) - F(x_{i-1}, y_1), \quad i \geq 2$, and $f(x_i, y_j) = F(x_i, y_j) - F(x_i, y_{j-1}) - F(x_{i-1}, y_j) + F(x_{i-1}, y_{j-1}), \quad i, j \geq 2$.*

Theorem 2.6 *Let $f(x_1, ..., x_n)$ and $F(x_1, ..., x_n)$ denote the joint pdf and cdf for a continuous multivariate random variable $\boldsymbol{X} = (X_1, ..., X_n)$. Then the joint pdf for $\boldsymbol{X}$ is obtained as $f(x_1, ..., x_n) = \frac{\partial^n F(x_1, ..., x_n)}{\partial x_1 \cdots \partial x_n}$ wherever $f(\cdot)$ is continuous and 0 elsewhere.*

Theorem 2.7 *Let $\boldsymbol{X} = (X_1, X_2)$ be a discrete random variable with joint pdf $f(x_1, x_2)$ and a range $R(\boldsymbol{X}) = R(X_1) \times R(X_2)$. The marginal pdfs are given by $f_1(x_1) = \sum_{x_2 \in R(X_2)} f(x_1, x_2)$ and $f_2(x_2) = \sum_{x_1 \in R(X_1)} f(x_1, x_2)$.*

Theorem 2.8 *Let $\boldsymbol{X} = (X_1, X_2)$ be a continuous random variable with joint pdf $f(x_1, x_2)$. The corresponding marginal pdfs are given by $f_1(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_2$ and $f_2(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_1$.*

Theorem 2.9 *The random variables X_1 and X_2 with joint pdf $f(x_1, x_2)$ and marginal pdfs $f_1(x_1)$ and $f_2(x_2)$ are independent, iff $f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2) \quad \forall (x_1, x_2)$ (except possibly at points of discontinuity for a joint continuous pdf f).*

Theorem 2.10 *The random variables $X_1, \dots, X_n$ with joint pdf $f(x_1, \dots, x_n)$ and marginal pdfs $f_i(x_i)$, $i = 1, \dots, n$, are all independent of each other, iff $f(x_1, \dots, x_n) = \prod_{i=1}^n f_i(x_i) \quad \forall (x_1, \dots, x_n)$ (except possibly at points of discontinuity for a joint continuous pdf f).*

Theorem 2.11 *If X_1 and X_2 are independent random variables, and if Y_1 and Y_2 are defined as functions $Y_1 = g_1(X_1)$ and $Y_2 = g_2(X_2)$, then Y_1 and Y_2 are independent.*

Theorem 2.12 (Change of Variables (Univariate Case)) *Let X be a continuous random variable with a pdf $f(x)$ with support $\Xi = \{x : f(x) > 0\}$. Suppose that $y = g(x)$ is a continuously differentiable function with $\frac{dg(x)}{dx} \neq 0 \forall x$ in some open interval Δ containing Ξ, and an inverse $x = g^{-1}(y)$ defined $\forall\ y \in \Psi = \{y : y = g(x), x \in \Xi\}$. Then the pdf of $Y = g(X)$ is given by $h(y) = f\left(g^{-1}(y)\right) \left|\frac{dg^{-1}(y)}{dy}\right| \quad$ for $\quad y \in \Psi$.*

Theorem 2.13 (Change of Variables (Multivariate Case)) *Let $\boldsymbol{X}$ be a continuous $(n \times 1)$ random vector with joint pdf $f(\boldsymbol{x})$ with support Ξ. Furthermore, let $\boldsymbol{g}(\boldsymbol{x})$ be a $(n \times 1)$ vector function which is continuously differentiable $\forall\ \boldsymbol{x}$ in some open rectangle, Δ, containing Ξ, and with an inverse $\boldsymbol{x} = \boldsymbol{g}^{-1}(\boldsymbol{y})$, which exists $\forall\ \boldsymbol{y} \in \boldsymbol{\Psi} = \{\boldsymbol{y} : \boldsymbol{y} = \boldsymbol{g}(\boldsymbol{x}), \boldsymbol{x} \in \Xi\}$. Assume that the Jacobian matrix $\boldsymbol{J}$ satisfies $det(\boldsymbol{J}) \neq 0$, and that all partial derivatives in $\boldsymbol{J}$ are continuous $\forall\ \boldsymbol{y} \in \boldsymbol{\Psi}$. Then the joint pdf of $\boldsymbol{Y} = \boldsymbol{g}(\boldsymbol{x})$ is given by $\quad h(\boldsymbol{y}) = f\left(g_1^{-1}(\boldsymbol{y}), \dots, g_n^{-1}(\boldsymbol{y})\right) |det(\boldsymbol{J})| \quad$ for $\quad \boldsymbol{y} \in \boldsymbol{\Psi}$.*

3. Moments of random variables

Theorem 3.1 *If $|x| < c\ \forall\ x \in R(X)$, for some choice of $c \in (0, \infty)$. Then $E(X)$ exists.*

Theorem 3.2 *Let X be a random variable with pdf $f(x)$. Then the expectation of random variable $Y = g(X)$ is given by $E(g(X)) = \sum_{x \in R(X)} g(x) f(x)$ (discrete) or $E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx$ (continuous).*

Theorem 3.3 (Jensen's Inequality) *Let X be a non-degenerate random variable with expectation $E(X)$, and let g be a function with smooth derivative on an open interval I containing $R(X)$ (that is $R(X) \subseteq I$). If g is convex on I, then $E(g(X)) \geq g(E(X))$; If g is strictly convex on I, then $E(g(X)) > g(E(X))$.*

Theorem 3.4 *If c is a constant, then $E(c) = c$.*

Theorem 3.5 *If c is a constant, then $E(cX) = cE(X)$.*

Theorem 3.6 $\mathrm{E}(\sum_{i=1}^k g_i(X)) = \sum_{i=1}^k E(g_i(X))$.

Corollary 3.1 $E(a+bX)=a+bE(X)$.

Theorem 3.7 *Let* $(X_1,...,X_n)'$ *be a multivariate random variable with joint pdf* $f(x_1,...,x_n)$. *The expectation of* $Y = g(X_1,...,X_n)$ *is* $E(Y) = \underset{(x_1,...,x_n)\in R(X)}{\sum\cdots\sum} g(x_1,...,x_n)\, f(x_1,...,x_n)$ *(discrete case) and* $E(Y) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(x_1,...,x_n)\, f(x_1,...,x_n$ *(continuous case).*

Theorem 3.8 $E(\sum_{i=1}^{k} g_i(X_1,\ldots,X_n)) = \sum_{i=1}^{k} E(g_i(X_1,\ldots,X_n))$.

Corollary 3.2 $E(\sum_{i=1}^{k} X_i) = \sum_{i=1}^{k} E(X_i)$ *(Expectation of a sum is the sum of the expectations).*

Theorem 3.9 *Let* $X_1,\ldots X_n$ *be independent random variables. Then* $E(\prod_{i=1}^{n} X_i) = \prod_{i=1}^{n} E(X_i)$.

Theorem 3.10 (Law of Iterated Expectations) $E[E(g(Y)|X)] = E(g(Y))$.

Theorem 3.11 (Markov's inequality) *Let* X *be a random variable with pdf* f, *and let* g *be a nonnegative function of* X. *Then* $P(g(x)\geq a) \leq \frac{E(g(X))}{a}$ *for any* $a>0$.

Corollary 3.3 (Chebyshev's inequality) $P(|x-\mu|\geq k\sigma) \leq \frac{1}{k^2}$ *for* $k>0$.

Theorem 3.12 *If* $\mathrm{E}(|X|^r)$ *exists for an* $r>0$, *then* $\mathrm{E}(|X|^s)$ *exists* $\forall s\in[0,r]$.

Theorem 3.13 *If* $\mathrm{E}(|Y-\mu|^r)$ *exists for an* $r>0$, *then* $\mathrm{E}(|Y-\mu|^s)$ *exists* $\forall s\in[0,r]$.

Theorem 3.14 *Let* X *be a random variable for which the MGF* $M_X(t)$ *exists. Then* $\mu_r' = E(X^r) = \left.\frac{d^r M_X(t)}{dt^r}\right|_{t=0}$.

Theorem 3.15 (MGF Uniqueness Theorem) *If an MGF exists for a random variable* X *having pdf* $f(x)$, *then the MGF is unique; and, conversely, the MGF determines the pdf of* X *uniquely, at least up to a set of points having probability 0.*

Theorem 3.16 (Cauchy-Schwarz Inequality) $E(WZ)^2 \leq E(W^2)E(Z^2)$.

Theorem 3.17 (Covariance bound) $|\sigma_{XY}| \leq \sigma_X\sigma_Y$.

Theorem 3.18 (Correlation bound) $-1\leq \rho_{XY}\leq 1$.

Theorem 3.19 *If* X *and* Y *are independent, then* $\sigma_{XY}=0$ *and* $\rho_{XY}=0$.

Theorem 3.20 *If* $\rho_{XY}=1$ *or* -1, *then* $P(y=a+bx)=1$, *where* $b\neq 0$.

Theorem 3.21 *Let* $Y=\sum_{i=1}^{n} a_iX_i$, *where the* a_is *are real constants. Then* $E(Y)=\sum_{i=1}^{n} a_iE(X_i)$.

Theorem 3.22 *Let* $Y=\sum_{i=1}^{n} a_iX_i$, *where the* a_is *are real constants. Then* $\sigma_Y^2 = \sum_{i=1}^{n} a_i^2\sigma_{X_i}^2 + 2\sum_{i<j} a_ia_j\sigma_{X_iX_j}$.

Theorem 3.23 *Let* $\boldsymbol{Y}=\boldsymbol{A}\boldsymbol{X}$, *where* $\boldsymbol{A}=(a_{hm})$ *is a* $k\times n$ *matrix of real constants, and* $\boldsymbol{X}=(X_i)$ *is an* $n\times 1$ *vector of random variables. Then* $E(\boldsymbol{Y})=\boldsymbol{A}E(\boldsymbol{X})$.

Theorem 3.24 *Let* $\boldsymbol{Y}=\boldsymbol{A}\boldsymbol{X}$, *where* $\boldsymbol{A}=(a_{hm})$ *is a* $k\times n$ *matrix of real constants, and* $\boldsymbol{X}=(X_i)$ *is an* $n\times 1$ *vector of random variables. Then* $\mathbf{Cov}(\boldsymbol{Y})=\boldsymbol{A}\mathbf{Cov}(\boldsymbol{X})\boldsymbol{A}'$.

4. Parametric families of density functions

Discrete uniform $f(x;N)=\frac{1}{N}\mathbb{I}_{\{1,2,\ldots,N\}}(x)$ with $N\in\Omega=\{N: N \text{ is a positive integer}\}$; $\mu=\frac{N+1}{2}$, $\sigma^2=\frac{N^2-1}{12}$, $\mu_3=0$ and $M_X(t)=\frac{1}{N}\sum_{j=1}^{N} e^{jt}$.

Bernoulli $f(x;p)=p^x(1-p)^{1-x}\mathbb{I}_{\{0,1\}}(x)$ with $p\in\Omega=\{p:0\leq p\leq 1\}$; $\mu=p$, $\sigma^2=p(1-p)$, $\mu_3=2p^3-3p^2+p$ and $M_X(t)=pe^t+(1-p)$.

Binomial $f(x;n,p)=\dfrac{n!}{x!(n-x)!}p^x(1-p)^{n-x}$, $x\in\mathbb{N}$ and 0 otherwise, with $(n,p)\in\Omega=\{(n,p): n\in\mathbb{N}\setminus\{0\}, 0\leq p\leq 1\}$; $\mu=np$, $\sigma^2=np(1-p)$, $\mu_3=np(1-p)(1-2p)$ and $M_X(t)=(1-p+pe^t)^n$.

Multinomial $f(x_1,\dots,x_m;n,p_1,\dots,p_m)=\dfrac{n!}{\prod_{i=1}^m x_i!}\prod_{i=1}^m p_i^{x_i}$, $x_i=0,\dots,n$, $\sum_{i=1}^m x_i=n$ and 0 otherwise, with $(n,p_i)\in\Omega$ where $\Omega=\{(n,p_1,\dots p_m):n>0 \text{ integer}, 0\le p_i\le 1, \forall\, i, \sum_{i=1}^m p_i=1\}$; $\mu_i=np_i$, $\sigma_i^2=np_i(1-p_i)$, $\mathrm{Cov}(X_i,X_j)=-np_ip_j$, $\mu_{3,i}=np_i(1-p_i)(1-2p_i)$ and $M_X(t)=\left(\sum_{i=1}^{m-1}p_ie^{t_i}+p_m\right)^n$.

Negative binomial (Pascal) $f(x;r,p)=\dbinom{x-1}{r-1}p^r(1-p)^{x-r}$, $x=r,r+1,\dots$ and 0 otherwise, with $(r,p)\in\Omega$ where $\Omega=\{(r,p):r>0\ \text{ integer}, 0<p<1\}$; $\mu=\frac{r}{p}$, $\sigma^2=\frac{r(1-p)}{p^2}$, $\mu_3=\frac{r\left((1-p)+(1-p)^2\right)}{p^3}$ and, for $t<-\ln(1-p)$, $M_X(t)=e^{rt}p^r\left(1-(1-p)e^t\right)^{-r}$.

Poisson $f(x;\lambda)=\frac{e^{-\lambda}\lambda^x}{x!}$, for $x=0,1,2...$ and 0 otherwise, with $\lambda\in\Omega=\{\lambda:\lambda>0\}$; $\mu=\lambda,\sigma^2=\lambda,\mu_3=\lambda$ and $M_X(t)=e^{\lambda\left(e^t-1\right)}$.

Theorem 4.1 *Let X be the number of times a certain event occurs in the interval $[0,t]$. If the experiment underlying X follows a Poisson process, then the pdf of X is the Poisson density.*

Hypergeometric $f(x;M,K,n)=\frac{\binom{K}{x}\binom{M-K}{n-x}}{\binom{M}{n}}$ for integer values $\max[0,n-(M-K)]\le x\le\min(n,K)$ and 0 otherwise, $(M,K,n)\in\Omega=\{(M,K,n):M=1,2,\dots;\ K=0,1,\dots,M;\ n=1,2,\dots,M\}$; $\mu=\frac{nK}{M}$, $\sigma^2=n\left(\frac{K}{M}\right)\left(\frac{M-K}{M}\right)\left(\frac{M-n}{M-1}\right)$, $\mu_3=n\left(\frac{K}{M}\right)\left(\frac{MK}{M}\right)\left(\frac{M-2K}{M}\right)\left(\frac{M-n}{M-1}\right)\left(\frac{M-2n}{M-2}\right)$ and $M_X(t)=\frac{((M-n)!(M-K)!)}{(M-K-n)!M!}H\left(-n,-K,M-K-n+1,e^t\right)$, where $H(\cdot)$ is the hypergeometric function, $H(\alpha,\beta,r,Z)=1+\frac{\alpha\beta}{r}\frac{Z}{1!}+\frac{\alpha\beta(a+1)(\beta+1)}{r(r+1)}\frac{Z^2}{2!}+\cdots$.

Continuous uniform $f(x;a,b)=\frac{1}{b-a}\mathbb{I}_{[a,b]}(x)$ with $(a,b)\in\Omega=\{(a,b):-\infty<a<b<\infty\}$; $\mu=(a+b)/2$, $\sigma^2=(b-a)^2/12$, $\mu_3=0$ and $M_X(t)=\dfrac{e^{bt}-e^{at}}{(b-a)t}$ for $t\neq 0$ and $M_X(t)=1$ for $t=0$.

Gamma $f(x;\alpha,\beta)=\frac{1}{(\beta^\alpha\Gamma(\alpha))}x^{\alpha-1}e^{-x/\beta}\mathbb{I}_{(0,\infty)}(x)$ with $(\alpha,\beta)\in\Omega=\{(\alpha,\beta):\alpha>0,\beta>0\}$ and $\Gamma(\alpha)=\int_0^\infty y^{\alpha-1}e^{-y}dy$; $\mu=\alpha\beta,\sigma^2=\alpha\beta^2,\mu_3=2\alpha\beta^3$ and $M_X(t)=(1-\beta t)^{-\alpha}$ for $t<\beta^{-1}$, where $\Gamma(\alpha)=\int_0^\infty y^{\alpha-1}e^{-y}dy=(\alpha-1)!$. Also, $\Gamma(\alpha+1)=\alpha\Gamma(\alpha)$ for $\alpha>0$.

Theorem 4.2 *Let $X_1,...,X_n$ be independent random variables with $X_i\sim$ Gamma(α_i,β), $i=1,...,n$. Then $Y=\sum_{i=1}^n X_i\sim$ Gamma$(\sum_{i=1}^n\alpha_i,\beta)$.*

Theorem 4.3 *Let $X\sim$ Gamma(α,β). Then, for any $c>0$, $Y=cX\sim$ Gamma$(\alpha,\beta c)$.*

Exponential $f(x;\theta)=\frac{1}{\theta}e^{-x/\theta}\mathbb{I}_{(0,\infty)}(x)$ with $\theta\in\Omega=\{\theta:\theta>0\}$; $\mu=\theta,\sigma^2=\theta^2,\mu_3=2\theta^3$ and $M_X(t)=(1-\theta t)^{-1}$ for $t<\theta^{-1}$.

Theorem 4.4 *If $X\sim$ Exponential(θ), then $P(x>s+t|x>s)=P(x>t)\ \forall\ (t,s)>0$.*

Chi square $f(x;v)=\frac{1}{2^{v/2}\Gamma(v/2)}x^{(v/2)-1}e^{-x/2}\mathbb{I}_{(0,\infty)}(x)$ with $v\in\Omega=\{v:v\text{ is a positive integer}\}$degrees of freedom; $\mu=v,\sigma^2=2v,\mu_3=8v$ and $M_X(t)=(1-2t)^{-v/2}$ for $t<\frac{1}{2}$

Beta $f(x;\alpha,\beta)=\frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}\mathbb{I}_{(0,1)}(x)$ with $(\alpha,\beta)\in\Omega=\{(\alpha,\beta):\alpha>0,\beta>0\}$; $\mu=\frac{\alpha}{\alpha+\beta}$, $\sigma^2=\frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$, $\mu_3=\frac{2(\beta-\alpha)(\alpha\beta)}{(\alpha+\beta+2)(\alpha+\beta+1)(\alpha+\beta)^3}$ and $M_X(t)=\sum_{r=0}^\infty\frac{B(r+\alpha,\beta)}{B(\alpha,\beta)}\frac{t^r}{r!}$, where $B(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.

Univariate normal $f(x;\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$ with $(\mu,\sigma)\in\Omega=\{(\mu,\sigma):\mu\in(-\infty,\infty),\sigma>0\}$; $\mathrm{E}(X)=\mu$, $\mathrm{var}(X)=\sigma^2$, $\mu_3=0$ and $M_X(t)=\exp\left\{\mu t+\frac{1}{2}\sigma^2t^2\right\}$.

Theorem 4.5 *If $X\sim N(\mu,\sigma^2)$, then $Z=(X-\mu)/\sigma\sim N(0,1)$.*

Theorem 4.6 *If $X\sim N(0,1)$, then $Y=X^2\sim\chi^2_{(1)}$.*

Theorem 4.7 *Let $(X_1,\dots,X_n)$ independent $N(0,1)$-distributed random variables. Then $Y=\sum_{i=1}^n X_i^2\sim\chi^2_{(n)}$.*

Theorem 4.8 (Student-t Density) *Let $Z\sim N(0,1)$, let $Y\sim\chi^2_\nu$, and let Z and Y be independent. Then $T=\frac{Z}{\sqrt{Y/\nu}}$ has the t-density with ν degrees of freedom defined as $f(t;\nu)=\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}}\left(1+\frac{t^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)}\mathcal{I}_{(-\infty,\infty)}(t)$.*

Theorem 4.9 (F-DISTRIBUTION) *Let $Y_1 \sim \chi^2_{\nu_1}$, let $Y_2 \sim \chi^2_{\nu_2}$, and let Y_1 and Y_2 be independent. Then $F = \frac{Y_1/\nu_1}{Y_2/\nu_2}$ has the F-density with $\nu_1,\ \nu_2$ degrees of freedom defined as* $f(x;\nu_1,\nu_2) = \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)}\left(\frac{\nu_1}{\nu_2}\right) x^{\frac{\nu_1}{2}-1}\left(1+\frac{\nu_1}{\nu_2}x\right)^{-0.5(\nu_1+\nu_2)}$.

Multivariate normal $f(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^n|\boldsymbol{\Sigma}|}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}$, $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_n)'$, $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \sigma_n^2 \end{pmatrix}$,
$(\boldsymbol{\mu},\boldsymbol{\Sigma}) \in \Omega = \{(\boldsymbol{\mu},\boldsymbol{\Sigma}) : \boldsymbol{\mu}\in\mathbb{R}^n,\ \boldsymbol{\Sigma}$ is a $(n\times n)$p.d. symmetric matrix$\}$; $\mathrm{E}\boldsymbol{x} = \boldsymbol{\mu}$, $\mathrm{Cov}(\boldsymbol{x}) = \boldsymbol{\Sigma}$, $\underset{(n\times 1)}{\boldsymbol{\mu}_3} = [\mathbf{0}]$ and $M_{\boldsymbol{x}}(\boldsymbol{t}) = \exp\{\boldsymbol{\mu}'\boldsymbol{t} + (1/2)\,\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}\}$, $\boldsymbol{t} = (t_1,\ldots,t_n)'$.

Theorem 4.10 *Let $\boldsymbol{X}$ be an n-dimensional $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$-distributed random variable. Let $\boldsymbol{A}$ be any $(k\times n)$ matrix of constants with $rk(\boldsymbol{A}) = k$, and let $\boldsymbol{b}$ be any $(k\times 1)$ vector of constants. Then the $(k\times 1)$ random vector $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}$ is $N(\boldsymbol{A}\boldsymbol{\mu}+\boldsymbol{b}\ ,\ \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}')$ distributed.*

Theorem 4.11 *Let $\boldsymbol{Z}$ be an n-dimensional $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$-distributed random variable, where $\boldsymbol{Z} = \left[\boldsymbol{Z}'_{(1)},\boldsymbol{Z}'_{(2)}\right]'$, $\boldsymbol{\mu} = \left[\boldsymbol{\mu}'_{(1)},\boldsymbol{\mu}'_{(2)}\right]'$ and $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix}$ partitioned conformably. Then the marginal pdf of $\boldsymbol{Z}_{(1)}$ is $N(\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{11})$, and the marginal PDF of $\boldsymbol{Z}_{(2)}$ is $N(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_{22})$.*

Theorem 4.12 *Let $\boldsymbol{Z}$ be an n-dimensional $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$-distributed random variable, where$\boldsymbol{Z} = \left[\boldsymbol{Z}'_{(1)},\boldsymbol{Z}'_{(2)}\right]'$, $\boldsymbol{\mu} = \left[\boldsymbol{\mu}'_{(1)},\boldsymbol{\mu}'_{(2)}\right]'$ and $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix}$ partitioned conformably, and let $\boldsymbol{z}^0$ be an n-dimensional vector of constants partitioned conformably with the partition $\boldsymbol{Z}$ into $\boldsymbol{z}^0_{(1)}$ and $\boldsymbol{z}^0_{(2)}$. Then the conditional distributions of $\boldsymbol{Z}_{(1)}|\boldsymbol{Z}_{(2)} = \boldsymbol{z}^0_{(2)}$ and $\boldsymbol{Z}_{(2)}|\boldsymbol{Z}_{(1)} = \boldsymbol{z}^0_{(1)}$ are*

$$\begin{aligned}
\boldsymbol{Z}_{(1)}|(\boldsymbol{Z}_{(2)} = \boldsymbol{z}^0_{(2)}) &\sim N\left(\boldsymbol{\mu}_{(1)} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\left[\boldsymbol{z}^0_{(2)} - \boldsymbol{\mu}_{(2)}\right]\ ,\ \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\right) \\
\boldsymbol{Z}_{(2)}|(\boldsymbol{Z}_{(1)} = \boldsymbol{z}^0_{(1)}) &\sim N\left(\boldsymbol{\mu}_{(2)} + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\left[\boldsymbol{z}^0_{(1)} - \boldsymbol{\mu}_{(1)}\right]\ ,\ \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\right).
\end{aligned}$$

Theorem 4.13 *Let $\boldsymbol{x} = (X_1,\ldots,X_n)'$ be a $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$-distributed random variable. Then $(X_1,\ldots,X_n)$ are independent iff $\boldsymbol{\Sigma}$ is a diagonal matrix with all covariances being zero.*

Exponential class $f(\boldsymbol{x};\boldsymbol{\Theta}) = \exp\left\{\sum_{i=1}^k c_i(\boldsymbol{\Theta})\,g_i(\boldsymbol{x}) + d(\boldsymbol{\Theta}) + z(\boldsymbol{x})\right\}$, $\boldsymbol{x}\in A$, and 0 otherwise, with $\boldsymbol{x} = (x_1,\ldots,x_n)'$ and parameters $\boldsymbol{\Theta} = (\Theta_1,\ldots,\Theta_k)'$; $c_i(\boldsymbol{\Theta})$, $d(\boldsymbol{\Theta})$ are real-valued functions of $\boldsymbol{\Theta}$ not depending on $\boldsymbol{x}$, $g_i(\boldsymbol{x})$, $z(\boldsymbol{x})$ are real-valued functions of $\boldsymbol{x}$ not depending on $\boldsymbol{\Theta}$ and $A\subset\mathbb{R}^n$ is a range/support which does not depend on $\boldsymbol{\Theta}$.

5. Basic asymptotics

Theorem 5.1 *Let $\{Y_n\}$ be a sequence of random variables having an associated sequence of MGFs $\{M_{Y_n}(t)\}$. Let $M_Y(t)$ be the MGF of Y. Then $Y_n \xrightarrow{d} Y$ iff $M_{Y_n}(t)\to M_Y(t)\quad \forall t\in(-h,h)$, for some $h>0$.*

Theorem 5.2 *Let $X_n \xrightarrow{d} X$, and let $g(X_n)$ be a continuous function which depends on n only via X_n. Then $g(X_n)\xrightarrow{d} g(X)$.*

Theorem 5.3 *Let $X_n \xrightarrow{p} X$, and let $g(X_n)$ be a continuous function which depends on n only via X_n. Then $plim\, g(X_n) = g(plim\, X_n) = g(X)$.*

Theorem 5.4 *For the sequences of random variables X_n, Y_n, and the constant a, it holds 1. plim $(aX_n) = a$ (plim X_n); 2. plim (X_n+Y_n) = plim X_n+plim Y_n (the plim of a sum = the sum of the plims); 3. plim (X_nY_n) = plim X_n plim Y_n (the plim of a product = the product of the plims); 4. plim (X_n/Y_n) = (plim X_n)/(plim Y_n) if $Y_n\neq 0$ and plim $Y_n\neq 0$.*

Theorem 5.5 $Y_n \xrightarrow{p} Y \Rightarrow Y_n \xrightarrow{d} Y$.

Theorem 5.6 $Y_n \xrightarrow{d} c \Rightarrow Y_n \xrightarrow{p} c$.

Theorem 5.7 (SLUTSKY'S THEOREMS) *Let $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} c$. Then, 1. $X_n+Y_n \xrightarrow{d} X+c$; 2. $X_n\cdot Y_n \xrightarrow{d} X\cdot c$; 3. $X_n/Y_n \xrightarrow{d} X/c$ if $Y_n\neq 0$ with probability 1 and $c\neq 0$.*

Theorem 5.8 $Y_n \overset{m}{\to} Y$ *iff* *1.* $E(Y_n) \to E(Y)$, *2.* $Var(Y_n) \to Var(Y)$, *3.* $Cov(Y_n, Y) \to Var(Y)$.

Corollary 5.1 $Y_n \overset{m}{\to} c$ *iff* $E(Y_n) \to c$ *and* $Var(Y_n) \to 0$.

Theorem 5.9 $Y_n \overset{m}{\to} Y \Rightarrow Y_n \overset{p}{\to} Y$.

Theorem 5.10 (Khinchin's WLLN) *Let* $\{X_n\}$ *be a sequence of iid random variables with finite expectations* $E(X_i) = \mu \; \forall i$. *Then* $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i \overset{p}{\to} \mu$.

Theorem 5.11 *Let* $\{X_n\}$ *be a sequence of random variables with finite variances, and let* $\{\mu_n\}$ *be the corresponding sequence of their expectations, Then* $\quad \bar{X}_n - \bar{\mu}_n \overset{p}{\to} 0 \quad$ *iff* $\quad E\left[\frac{(\bar{X}_n-\bar{\mu}_n)^2}{1+(\bar{X}_n-\bar{\mu}_n)^2}\right] \to 0$.

Theorem 5.12 *Let* $\{X_n\}$ *be a sequence of random variables with respective expectations given by* $\{\mu_n\}$. *If* $Var(\bar{X}_n) \to 0$, *then* $\bar{X}_n - \bar{\mu}_n \overset{p}{\to} 0$.

Theorem 5.13 (Lindeberg-Lévy) *Let* $\{X_n\}$ *be a sequence of iid random variables with* $E(X_i) = \mu$ *and* $Var(X_i) = \sigma^2 \in (0, \infty) \; \forall i$. *Then* $Y_n = \frac{1}{\sqrt{n}\sigma}\left(\sum_{i=1}^n X_i - n\mu\right) = \frac{\sqrt{n}(\bar{X}_n-\mu)}{\sigma} \overset{d}{\to} N(0,1)$.

Theorem 5.14 (Lindeberg's CLT) *Let* $\{X_n\}$ *be a sequence of independent random variables with* $E(X_i) = \mu_i$ *and* $Var(X_i) = \sigma_i^2 < \infty \, \forall \, i$. *Define* $b_n^2 = \sum_{i=1}^n \sigma_i^2$, $\bar{\sigma}_n^2 = n^{-1}\sum_{i=1}^n \sigma_i^2$, $\bar{\mu}_n = n^{-1}\sum_{i=1}^n \mu_i$, *and let* f_i *be the PDF of* X_i. *If* $\forall \, \varepsilon > 0$, $\lim_{n\to\infty} \frac{1}{b_n^2}\sum_{i=1}^n \int_{(x_i-\mu_i)^2 \ge \varepsilon b_n^2} (x_i-\mu_i)^2 f_i(x_i)\, dx_i = 0$ *(continuous case) or* $\lim_{n\to\infty} \frac{1}{b_n^2}\sum_{i=1}^n \sum_{\substack{(x_i-\mu_i)^2 \ge \varepsilon b_n^2 \\ f_i(x_i)>0}} (x_i-\mu_i)^2 f_i(x_i) = 0$ *(discrete case), then* $\quad \frac{\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i}{\left(\sum_{i=1}^n \sigma_i^2\right)^{1/2}} = \frac{n^{1/2}\left(\bar{X}_n-\bar{\mu}_n\right)}{\bar{\sigma}_n} \overset{d}{\to} N(0,1)$.

Theorem 5.15 (CLT for bounded variables) *Let* $\{X_n\}$ *be a sequence of independent random variables such that* $P(|x_i| \le m) = 1 \; \forall \; i$ *for some* $m \in (0,\infty)$, *and suppose* $E(X_i) = \mu_i$ *and* $Var(X_i) = \sigma_i^2 < \infty \, \forall \, i$. *If* $\sum_{i=1}^n Var(X_i) = \sum_{i=1}^n \sigma_i^2 \to \infty$ *as* $n \to \infty$, *then* $\frac{n^{1/2}\left(\bar{X}_n-\bar{\mu}_n\right)}{\bar{\sigma}_n} \overset{d}{\to} N(0,1)$.

Theorem 5.16 (Cramér-Wold Device) *The sequence of* $(k\times 1)$*-dim. random vectors* $\{\boldsymbol{X}_n\}$ *converges in distribution to the* $(k\times 1)$*-dim. random vector* $\boldsymbol{X}$ *iff* $\boldsymbol{\ell}'\boldsymbol{X}_n \overset{d}{\to} \boldsymbol{\ell}'\boldsymbol{X} \quad \forall \; \boldsymbol{\ell} \in \mathbb{R}^k$.

Corollary 5.2 $\boldsymbol{X}_n \overset{d}{\to} \boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *iff* $\boldsymbol{\ell}'\boldsymbol{X}_n \overset{d}{\to} \boldsymbol{\ell}'\boldsymbol{X} \sim N(\boldsymbol{\ell}'\boldsymbol{\mu}\,,\, \boldsymbol{\ell}'\boldsymbol{\Sigma}\boldsymbol{\ell})$.

Theorem 5.17 (Multivariate Lindeberg-Lévy) *Let* $\{\boldsymbol{X}_n\}$ *be a sequence of iid* $(k\times 1)$ *random vectors with* $E(\boldsymbol{X}_i) = \boldsymbol{\mu}$ *and* $\boldsymbol{Cov}(\boldsymbol{X}_i) = \boldsymbol{\Sigma} \, \forall \, i$, *where* $\boldsymbol{\Sigma}$ *is a* $(k\times k)$ *positive definite matrix. Then* $\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i - \boldsymbol{\mu}\right) \overset{d}{\to} N(\boldsymbol{0}, \boldsymbol{\Sigma})$.

Theorem 5.18 (Asymptotic Distribution of $g(\boldsymbol{X}_n)$; Delta method) *Let* $\{\boldsymbol{X}_n\}$ *be a sequence of* $(k\times 1)$ *random vectors such that* $\sqrt{n}(\boldsymbol{X}_n - \boldsymbol{\mu}) \overset{d}{\to} \boldsymbol{Z} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$. *Let* $g(\boldsymbol{x})$ *be a function that has first-order partial derivatives in a neighborhood of the point* $\boldsymbol{x} = \boldsymbol{\mu}$ *that are continuous at* $\boldsymbol{\mu}$, *and suppose the gradient vector of* $g(\boldsymbol{x})$ *evaluated at* $\boldsymbol{x} = \boldsymbol{\mu}$, $\boldsymbol{G}_{(1\times k)} = [\partial g(\boldsymbol{\mu})/\partial x_1 \ldots \partial g(\boldsymbol{\mu})/\partial x_k]$, *is not the zero vector. Then* $\quad \sqrt{n}\left(g(\boldsymbol{X}_n) - g(\boldsymbol{\mu})\right) \overset{d}{\to} N(0, \boldsymbol{G}\boldsymbol{\Sigma}\boldsymbol{G}')$ *and* $g(\boldsymbol{X}_n) \overset{a}{\sim} N\left(g(\boldsymbol{\mu}), n^{-1}\boldsymbol{G}\boldsymbol{\Sigma}\boldsymbol{G}'\right)$.