§6 Erwartungswerte, Varianz und Median

$(\Omega, \mathcal{A}, P)$ W'raum, $X: \Omega \to \mathbb{R}$ ZV'e mit Verteilung P_X.

Frage: Weitere Kenngrößen von X!

Zur Vorbereitung für Erwartungswert:

6.1. Abstrakter Transformationssatz:

$$\begin{tikzcd} \Omega \arrow[r, bend left, "X"] \arrow[dr, "h\circ X"'] & \tilde{\Omega} \arrow[d, "h"] \\ & \overline{\mathbb{R}} \end{tikzcd}$$

$(\Omega, \mathcal{A}, P)$ W'raum, $(\tilde{\Omega}, \tilde{\mathcal{A}})$ Meßraum,

$X: \Omega \to \tilde{\Omega}$ meßbar und $h: \tilde{\Omega} \to \overline{\mathbb{R}}$ $\mathcal{B}(\overline{\mathbb{R}})$-meßbar.

Dann: $h \circ X \in \mathcal{L}^1(\Omega, \mathcal{A}, P) \iff h \in \mathcal{L}^1(\tilde{\Omega}, \tilde{\mathcal{A}}, P_X)$

In diesem Fall: $\int_\Omega h \circ X \, dP = \int_{\tilde{\Omega}} h \, dP_X$.

Beweis: ① $h = 1_A$ für $A \in \tilde{\mathcal{A}} \Rightarrow \int_{\tilde{\Omega}} 1_A \, dP_X = P_X(A) = P(X^{-1}(A)) = \int_\Omega \overbrace{1_{X^{-1}(A)}}^{=1_A \circ X} dP$

also gelten Aussage.

② Linearität der Integrale $\Rightarrow$ Aussage OK für $h \in E^+(\tilde{\Omega}, \tilde{\mathcal{A}})$.

③ $h \geq 0$ meßbar $\Rightarrow \exists (h_n)_{n \in \mathbb{N}} \subset E^+(\tilde{\Omega}, \tilde{\mathcal{A}})$ mit $h_n \uparrow h \Rightarrow h_n \circ X \uparrow h \circ X$

$\Rightarrow \int h \, dP_X \longleftarrow \int_{\tilde{\Omega}} h_n \, dP_X = \int_\Omega h_n \circ X \, dP \longrightarrow \int h \circ X \, dP \Rightarrow$ Beh.

④ Linearität $\Rightarrow$ allg. Fall ■

6.2. Bsp: $(\Omega, \mathcal{A}, P)$ W'raum, $X: \Omega \to \mathbb{R}$ ZV'e mit Verteilung P_X

① $\varphi: \mathbb{R} \to \mathbb{R}$ streng monoton wachsend. Frage: $P_{\varphi(X)} = ?$

Lösung über Verteilungsfkt.: $F_{\varphi(X)}(t) = P(\varphi(X) \leq t) = P(X \leq \varphi^{-1}(t)) = F_X(\varphi^{-1}(t))$

② Bsp.: F_X mit Dichte f_X und $\varphi \in \mathcal{C}^{(1)}(\mathbb{R})$ zusätzlich

$\Rightarrow P_{\varphi(X)}$ hat Dichte $f_{\varphi(X)}$ mit $f_{\varphi(X)} = \frac{d}{dt} F_{\varphi(X)}(t) \underset{\text{oben}}{=} f_X(\varphi^{-1}(t)) (\varphi^{-1})'(t)$.

③ Sei $n \in \mathbb{N}$ und $X: \Omega \to \mathbb{R}$ ZV'e mit $X^n \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$, d.h. $h(x) = x^n$

$\Rightarrow \int_\Omega X^n \, dP = \int_{\mathbb{R}} x^n \, dP_X(x)$

6.3. Def.: $(\Omega, \mathcal{A}, P)$ W'raum, $X: \Omega \to \mathbb{R}$ ZV'e

① Falls $X \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ ist (oder äquivalent: $\int_{\mathbb{R}} |x| \, dP_X(x) < \infty$), so sagt man: X hat einen Erwartungswert, und dieser ist

$$E(X) := \int_\Omega X\, dP \overset{6.2(3)}{=} \int_{\mathbb{R}} x\, dP_X(x) \in \mathbb{R}.$$

(2) Falls für $n \in \mathbb{N}$ $X^n \in L^1(\Omega, \mathcal{A}, P)$, so sagt man: X hat das n-te Moment $\mu_n(X) := E(X^n) = \int_{\mathbb{R}} x^n\, dP_X(x) \in \mathbb{R}$.

(3) $L^2(\Omega, \mathcal{A}, P) := \{X : \Omega \to \mathbb{R}$ ZV'e mit $\int_\Omega X^2\, dP < \infty\}$.
Raum der quadrat-integrablen ZV'en.

<u>6.4 Bsp.:</u>

(1) Sei $X : \Omega \to \mathbb{R}$ <u>diskrete ZV'e</u>, d.h. der Wertebereich $X(\Omega) = \{x_1, x_2, \dots\}$ ist höchstens abzählbar $\leadsto P_X = \sum_{k \geq 1} p_k \delta_{x_k}$, $\sum_{k \geq 1} p_k = 1$, $\forall p_k \geq 0$.
Dabei $p_k = P(X = x_k) = P_X(\{x_k\})$.
Dann: $E(X) = \int_{\mathbb{R}} x\, dP_X(x) \overset{\S 5}{=} \sum_{k \geq 1} p_k x_k$
und $E(X^n) = \sum_{k \geq 1} p_k x_k^n$ } falls existent.

(2) <u>Bsp:</u> X geometrisch verteilt mit Index $p \in]0,1[$, d.h. $P(X = k) = p(1-p)^{k-1}$ für $k \in \mathbb{N}$ $\Rightarrow E(X) = \sum_{k=1}^{\infty} p(1-p)^{k-1} \cdot k \overset{(*)}{=} 1/p$.
zu $(*)$: $\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k$ für $|x| < 1 \Rightarrow \frac{1}{(1-x)^2} = \sum_{k=1}^{\infty} k x^{k-1}$. setze $x = 1-p$!

(3) <u>Bsp.:</u> X $B_{n,p}$-verteilt $\Rightarrow E(X) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} =$

$$= \sum_{k=1}^{n} \frac{n\,(n-1)!}{((n-1)-(k-1))!\,(k-1)!}\, p\, p^{k-1} (1-p)^{(n-1)-(k-1)} = n \cdot p \underbrace{\sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k}}_{=1} = n \cdot p$$

(4) P_X habe Lebesgue-Dichte f $\Rightarrow$
$\mu_n(X) = E(X^n) = \int x^n\, dP_X(x) \overset{\S 5}{=} \int x^n f(x)\, dx$

(5) <u>Bsp.:</u> X exponentialverteilt mit Index $\lambda > 0$ $\Rightarrow$
$E(X) = \int_0^\infty x \cdot \lambda e^{-\lambda x}\, dx = \left(-x e^{-\lambda x} - \frac{1}{\lambda} e^{-\lambda x}\right)\Big|_0^\infty = 1/\lambda$

(6) X $N(\mu, \sigma^2)$-verteilt $\Rightarrow$

$$E(X) = \int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}\, dx \overset{y = x-\mu \text{ Subst.}}{=} \frac{1}{\sqrt{\dots}} \int_{-\infty}^{\infty} (y+\mu)\, e^{-y^2/2\sigma^2}\, dx = \mu$$

Teil $= 0$, weg. Punktsymm.!

6.5. Rechenregeln: $X,Y:\Omega\to\mathbb{R}$ ZV'e

① **Linearität:** $a,b\in\mathbb{R}$, $X,Y\in L^1(\Omega,\mathcal{A},P) \Rightarrow E(aX+bY)=a\,E(X)+b\,E(Y)$

(klar wg. Linearität d. Integrale!)

② $X,Y\in L^1(\Omega,\mathcal{A},P)$ und X,Y unabhängig $\Rightarrow X\cdot Y\in L^1(\Omega,\mathcal{A},P)$ und

$$E(XY)=E(X)\cdot E(Y).$$

Beachte: Aussage i.a. falsch, Unabhängig wichtig!

allg. Beweis von ② in §7; hier diskrete Fall:

$$\left.\begin{array}{l} X \text{ nimmt Werte } \{x_1,x_2,\dots\} \text{ an} \\ Y \text{ —"— } \{y_1,y_2,\dots\} \text{ an} \end{array}\right\} \Rightarrow E(XY)=\sum_{i,j} x_i y_j\, P(X=x_i \text{ und } Y=y_j)$$

$$=\sum_{i,j} x_i y_j\, P(X=x_i)\, P(Y=y_j) = E(X)E(Y)$$

③ Sei $m\le n\in\mathbb{N}$. Falls $\mu_n(X)$ existiert, d.h. $E(|X|^n)<\infty$, so existiert auch $\mu_m(X)$, denn:

$$\int_{\mathbb{R}}|x|^m\,dP_X(x)=\int_{-1}^{1}|x|^m\,dP_X(x)+\int_{\{x:|x|>1\}}|x|^m\,dP_X(x)\le 1+\int|x|^n\,dP_X(x)<\infty$$

6.6. Bsp.: $X_1,\dots,X_n$ Ergebnisse von n unabhängigen ZVen mit $P_{X_i}=p\,\delta_1+(1-p)\delta_0$, $i\in\mathbb{N}$.

$\Rightarrow X:=X_1+\dots+X_n$ $B_{n,p}$-verteilt und

$E(X_i)=0(1-p)+1\cdot p=p \Rightarrow E(X)=n\cdot E(X_1)=n\cdot p$. Einfacher als 6.4(3)!!

Merke: Für Berechnung von $E(X)$ Darstellung von X oft wichtig; vgl. Übungen!!

6.7 Cauchy-Schwarz-Ungleichung: $X,Y\in L^2(\Omega,\mathcal{A},P) \Rightarrow$

$E(XY)^2\le E(X^2)\,E(Y^2)$ und $X\cdot Y\in L^1(\Omega,\mathcal{A},P)$.

Beweis: Falls $E(Y^2)=0$, ist $P(Y=0)=1$ und $E(XY)=0 \Rightarrow$ Beh.

Sei nun $E(Y^2)>0$. Dann $0\le E\left(\left[E(Y^2)X-E(XY)Y\right]^2\right)=$

$E(Y^2)^2E(X^2)+\cancel{E(XY)^2E(Y^2)}-\cancel{2}E(XY)^2E(Y^2) \Rightarrow$ Beh. ∎

6.8. Varianz: $X,Y\in L^2(\Omega,\mathcal{A},P) \overset{6.7}{\Rightarrow} E(|XY|)<\infty$

① $\mathrm{Kov}(X,Y):=E\big((X-E(X))(Y-E(Y))\big)$ Kovarianz von X,Y.

② $\mathrm{Var}(X):=\mathrm{Kov}(X,X)$ Varianz von X

③ $\sigma(X) := Var(X)^{1/2} \geq 0$ Streuung od. Standardabweichung von X

④ $\rho_{XY} := Kov(X,Y) / \sigma(X)\cdot\sigma(Y)$ Korrelationskoeff.

⑤ X, Y unkorreliert $:\Leftrightarrow Kov(X,Y) = 0 \Leftrightarrow \rho_{X,Y} = 0$

6.9. Bem.:

① $Var X$, $\sigma(X)$ messen Abweichung von X vom „Mittelwert" $E(X)$.
$E(|X - E(X)|)$ mögl. Variante, aber math. schwieriger!

② $\rho_{X,Y}$ mißt Abhängigkeit von X und Y. Genauer:

- Immer $\rho_{X,Y} \in [-1, 1]$ wg. CSU.
- X, Y unabhängig $\Rightarrow X, Y$ unkorreliert, denn
$E((X-E(X))(Y-E(Y))) = E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y) = 0$
- $\rho_{X,Y} > 0 \Leftrightarrow: X, Y$ positiv korreliert (Bsp: Schuhgröße $\leftrightarrow$ Körpergröße)
- $\rho_{X,Y} < 0 \Leftrightarrow: X, Y$ negativ korreliert (Bsp: Zigaretten pro Tag u. Lebenserwartung)
- $\rho_{X,X} = +1$, $\rho_{X,-X} = -1$ Extremfälle

6.10. Rechenregeln: $X, Y, X_1, \ldots, X_n \in L^2(\Omega, \mathcal{A}, P) \Rightarrow$

① $Kov(X,Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y)$
Speziell: $Var(X) = E(X^2) - E(X)^2 \geq 0$

② $a, b \in \mathbb{R} \Rightarrow E(aX + b) = aE(X) + b$, $Var(aX+b) = a^2 Var(X)$
$\sigma(aX+b) = |a|\,\sigma(X)$.

③ $Kov(X,Y) = Kov(Y,X)$

④ $Var(X_1 + \ldots + X_n) = E((\sum_{i=1}^n X_i - E(X_i))^2) = \sum_{i,j} Kov(X_i, X_j)$

⑤ X, Y unabhängig $\Rightarrow Kov(X,Y) = 0$ (vgl. oben)

⑥ $X_1, \ldots, X_n$ unabhängig $\Rightarrow Var(X_1 + \ldots + X_n) = \sum_{i=1}^n Var(X_i)$ (wg. ④ + ⑤)
„Formel von Bienaymé"

6.11. Bsp.:

① X $B_{n,p}$-verteilt, $Var(X) = ?$
Lsg 1: $Var(X) = E(X^2) - E(X)^2 = \sum_{k=0}^n k^2 \binom{n}{k} p^k (1-p)^{n-k} - (np)^2 = ? \ldots$
Lsg 2: $X = X_1 + \ldots + X_n$ wie oben $\Rightarrow Var(X) = n \cdot Var(X_1) = n(1^2 p + 0^2(1-p) - p^2)$

$= np(1-p)$.

(2) X $\mathcal{N}(\mu,\sigma^2)$-verteilt $\Rightarrow$ $E(X)=\mu$, $Var(X)=\sigma^2$.

Beweis:

Fall $\mu=0, \sigma^2=1$: $Var(X) = E(X^2) - \overset{0}{E(X)^2} = \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \overset{\text{partiell}}{=}$ $\quad u(x)=x$, $v'(x) = x e^{-x^2/2}$

$= \frac{1}{\sqrt{2\pi}} \left[x(-e^{-x^2/2}) \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} -e^{-x^2/2} dx \right] = 1$

Allg. Fall durch Substitution auf Fall oben reduzierbar ■

Fazit: Berechnung von $E(X)$, $Var(X)$ oft schwierig, auch im diskreten → Tricks nötig.

Systematischer Ansatz für $\mathbb{N}_0$-wertige ZVe über:

6.12. Erzeugende Fkt'en: $X:\Omega \to \mathbb{N}_0 \subset \mathbb{R}$ ZVe mit $P(X=n)=p_n$ $(n\geq 0)$.

Die Erzeugende Fkt von X ist $g_X(t) := \sum_{n=1}^{\infty} p_n t^n$ für $t\in\mathbb{C}$, $|t|\leq 1$.

Dann:
(1) $g_X'(1) = E(X)$
(2) $g_X''(1) = E((X-1)X)$
(3) $Var(X) = g_X''(1) + g_X'(1)(1-g_X'(1))$.

Beweis: $g_X'(1) = \sum_n n p_n = E(X)$ und $g_X''(1) = \sum n(n-1)p_n = E((X-1)X)$

(3) folgt aus (1) + (2) + $Var(X) = E(X^2) - E(X)^2$. ■

6.13. Bsp.: X π_λ-verteilt $\Rightarrow g_X(t) = \sum_0^{\infty} t^n \frac{\lambda^n}{n!} e^{-\lambda} = e^{-\lambda} e^{\lambda t} = e^{\lambda(t-1)}$

$\Rightarrow E(X) = g_X'(1) = \lambda e^{-\lambda} e^{\lambda t}\Big|_{t=1} = \lambda$

$Var(X) = g_X''(1) + \lambda(1-\lambda) = \ldots = \lambda$ ■

6.14. Median:

Def.: $m\in\mathbb{R}$ heißt Median der ZVen $X:\Omega\to\mathbb{R}$, falls

$P(X\geq m) \geq 1/2$ und $P(X\leq m)\geq 1/2$.

Beachte: Median i.a. nicht eindeutig, Bsp.: $P_X = \frac{1}{2}(\delta_0+\delta_1)$.

Hier jedes $m\in[0,1]$ Median!!

Weiteres Problem: Mediane i.a. nicht formelmäßig bestimmbar

i.a. Bestimmung durch Intervallschachtelung.

Bsp: X symmetrisch um $E(X)$, d.h. $\forall c \geq 0$: $P(X \geq E(X)+c) = P(X \leq E(X)-c)$

Dann: $E(X)$ ist ein Median.

Bsp.: $N(\mu, \sigma^2)$ hat μ als einzigen Median.

Median math. unwichtig, aber evtl. wichtig für Interpretation.

Bsp.: Dorf A hat 100 Einwohner, 99 mit ∅-Einkommen 30 000,- sowie Bill G. mit 10^8,- Einkommen.

Dorf B hat 100 Einwohner, jeder mit durchschnittl. 100 000,-

	A	B
Erwartungswert Einkommen	$\approx 10^6$	10^5
Median	30 000	10^5

Was aussagekräftiger? Antwort kontextabhängig.

Steueraufkommen EW wichtiger, allg. Wohlstand Median besser.

Allg.: Median berücksichtigt Ausreißer weniger!

6.15. Satz. $X, Y: (\Omega, \mathcal{A}, P) \to \mathbb{R}$ ZV'e, $\varepsilon > 0$:

① Markov-Ungl.: $f: [0,\infty[\to [0,\infty[$ monoton wachsend $\Rightarrow$
$P(|Y| \geq \varepsilon) \leq E(f(|Y|))/f(\varepsilon)$.

② Tschebyschev-Ungl.: $E(X^2) < \infty \Rightarrow P(|X - E(X)| \geq \varepsilon) \leq Var(X)/\varepsilon^2$.

Beweis: ① f monoton $\Rightarrow$ f meßbar $\Rightarrow$ $Z := 1\!\!1_{\{|Y| \geq \varepsilon\}} f(\varepsilon)$ ZV'e mit $Z \leq f(|Y|)$
$\Rightarrow P(|Y| \geq \varepsilon) \cdot f(\varepsilon) = E(Z) \leq E(f(|Y|)) \Rightarrow$ Beh.

② Setze in ① $Y := X - E(X)$, $f(x) = x^2$; $x \geq 0$. ■

6.16. Schwaches Gesetz d. großen Zahlen:

$X_1, \dots, X_n \in L^2(\Omega, \mathcal{A}, P)$ unkorreliert mit gleichem Erwartungswert $E(X_i) =: m$
und: $\forall i = 1, \dots, n$: $Var(X_i) \leq M$. Sei $\overline{X}_n := \frac{1}{n}(X_1 + \dots + X_n)$.
$\Rightarrow \forall \varepsilon > 0$: $P(|\overline{X}_n - m| \geq \varepsilon) \leq M/\varepsilon^2 n \xrightarrow[n\to\infty]{} 0$.

Beweis: $E(\overline{X}_n) = m$, $Var(\overline{X}_n) = \frac{1}{n^2}(Var(X_1) + \dots + Var(X_n)) \leq M/n$

Aussage folgt aus T-Ungleichung ■

Bemerkung: $\overline{X}_n$ heißt empirisches Mittel aus $X_1, \dots, X_n$.

6.16 sagt: $\overline{X}_n \to m$ in schwachem Sinn.

Mehr über Konvergenzbegriffe und starkes Gesetz d. großen Zahlen später.

6.17 Bsp.: Wahlvorhersage:

Frage: Mindestumfang n einer Stichprobe für Vorhersage des Stimmanteils $p \in]0,1[$ einer Partei bei Wahl.

Forderung: W'keit für absoluten Fehler über 0,01 ($\hat{=}$ 1%) maximal 0,05.

Modell: n „unabhängige" Testpersonen, die Bevölkerung repräsentieren

$\leadsto$ $X_1, \dots, X_n$ unabhängige, $\{0,1\}$-wertige ZV'e mit gemeinsamer Verteilung $p\delta_1 + (1-p)\delta_0$.

$\leadsto$ $S_n := X_1 + \dots + X_n$ $B_{n,p}$-verteilt.

Schätzer für unbekannte p ist S_n/n.

Fehlerabschätzung mit T-Ungl.: $P\left(\left|S_n/n - p\right| \geq \varepsilon\right) \overset{6.16}{\leq} \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n(0{,}01)^2} \overset{!}{\leq} 0{,}05$

„$\leq$" erfüllt für $n \geq 50\,000$.

Frage: Kann man Ergebnis verbessern: Schärfere Ungleichung $\leadsto$ n kleiner wählbar.
$\leadsto$ später

6.18 Diskrete bedingte Erwartungswerte:

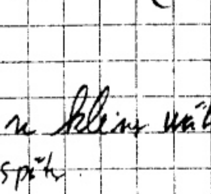
$(\Omega, \mathcal{A}, P)$ W'raum, $A \in \mathcal{A}$ mit $P(A) > 0$

Für $B \in \mathcal{A}$ ist $P(B|A) := \frac{1}{P(A)} P(B \cap A)$ bed. W'keit von B unter Bdg. A.

Für $X \in L^1(\Omega, \mathcal{A}, P)$ heißt $E(X|A) := \frac{1}{P(A)} \int_A X \, dP$ bed. Erwartung von X unter der Bdg. A.

Zusammenhang: $P(B|A) = E(1_B | A)$.

Also: Bed. W'keit auffassen als Spezialfall bed. Erwartungen.



Bsp.: $\Omega = \{(m,n) \in \mathbb{N}_0 : m + n \leq 2\}$

mit Gleichverteilung, $|\Omega| = 6$.

Übung: Polyasches Urnenmodell! $\hookrightarrow$

$X_i : \Omega \to \mathbb{R}$, $(n_1, n_2) \mapsto n_i$ i-te Koordinate.

Dann: $E(X_1) = \frac{3}{6} \cdot 0 + \frac{2}{6} \cdot 1 + \frac{1}{6} \cdot 2 = \frac{4}{6} = 2/3$.

$E(X_1 | X_2 = 1) = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 1 = 1/2$.

6.19. Nullmengen und fast sichere Ereignisse:

① Sei $(\Omega, \mathcal{A}, P)$ Maßraum. $A \in \mathcal{A}$ heißt P-Nullmenge, falls $P(A) = 0$

② Sei $(\Omega, \mathcal{A}, P)$ W'raum. $A \in \mathcal{A}$ heißt fast sicheres Ereignis, falls
$P(A) = 1$ $(\Leftrightarrow P(\Omega \setminus A) = 0)$

Sprechweise: Eine Aussage gilt auf $(\Omega, \mathcal{A}, P)$ fast sicher, falls
$$P(\{\omega \in \Omega : \text{Aussage gilt für } \omega\}) = 1$$

Bsp.: $X, Y : \Omega \to \mathbb{R}$ ZV'e. Dann:
$X = Y$ fast sicher (kurz: f.s.) $:\Leftrightarrow P(\{\omega \in \Omega : X(\omega) = Y(\omega)\}) = 1$.

③ Regeln:

ⓐ $(A_i)_{i \in \mathbb{N}} \subset \mathcal{A}$ Nullmengen $\Rightarrow \bigcup_{i \in \mathbb{N}} A_i$ Nullmenge
denn $P(\bigcup_i A_i) \leq \sum_{i \in \mathbb{N}} P(A_i) = 0$

ⓑ $(B_i)_{i \in \mathbb{N}} \subset \mathcal{A}$ f.s. Ereignisse $\Rightarrow \bigcap_{i \in \mathbb{N}} B_i$ f.s.