Fachhochschule
Münster University of
Applied Sciences

Kombinatorik und Wahrscheinlichkeit

Prof. Dr. Martin Poppe

Aufgaben zu Verteilungsfunktionen

1. Was zeichnet ein „Bernoulli Experiment" aus, und zu welcher Verteilungsfunktion führt ein solches Experiment?

2. Richtig/falsch Klausur
   Ein naiver Professor stellt als erste Klausuraufgabe 10 Thesen auf, von denen die Studenten notieren müssen, ob sie wahr oder falsch sind. Wenn nun ein Student überhaupt keine Ahnung hat, wie groß ist seine Chance, trotzdem mehr als 50% der Thesen korrekt zu beurteilen?

3. Multiple Choice Bogen
   Ein Student der Wirtschaftswissenschaften verirrt sich in eine Medizin-Klausur und findet 8 Fragen vor. Zu jeder Frage sind 3 Antworten vorgegeben, von denen eine oder zwei richtig sein können. Da er keine Ahnung von der Medizin hat, verteilt der ohne groß nachzudenken einzelne Kreuzchen und Kreuzchen-Paare. Wie groß ist die Wahrscheinlichkeit, für welche Zensur (8 richtig = 1, ...5 richtig = 4)?

4. Drunter und drüber
   Bitte berechnen Sie $\binom{n}{n}$, $\binom{n}{0}$, $\binom{n}{k} / \binom{n}{(n-k)}$ und $\binom{n}{1}$.

5. Der Honoratioren-Club eines Dorfes verkauft im November Glücks-Weihnachtskalender mit einer Gewinnchance von 4%. Wie groß ist die Wahrscheinlichkeit, dass der erste gekaufte Kalender ein Gewinn ist? Wie groß ist die Wahrscheinlichkeit, dass der 25. Kalender der erste Gewinn ist?

6. In einem Physik-Experiment wird nach seltenen Teilchen gesucht. Einer exotischen Theorie zur Folge entsteht dieses Teilchen bei jeder 10-millionsten Kernreaktion. Wie viele Reaktionen müssen analysiert werden, damit die die Theorie mit 95%-iger Wahrscheinlichkeit ausgeschlossen werden kann?

7. Im November ist die Wahrscheinlichkeit, das es in Münster an einem Tag regnet 80%. Wie wahrscheinlich ist es, dass vom ersten November an gerechnet die nächsten 24 Regentage genau bis zum 30. November fallen? Wie wahrscheinlich ist es, dass ein November genau 24 Regentage hat?

Aufgaben zu Varianzen und Erwartungswerten

8. In Steinfurt ist im Schnitt jedes 6. Jahr weiße Weihnacht. Wenn k die Anzahl der Jahre ist, die wir ab heute auf weiße Weihnacht warten müssen, was sind der Erwartungswert, die Varianz und die Standardabweichung von k????? Bitte schätzen sie: Wie oft passiert es in einem Jahrtausend, dass genau zwei weiße Weihnachten aufeinander folgen?????

9. Beim Würfeln sollte jede Zahl von 1 bis 6 die gleiche Wahrscheinlichkeit haben. Bitte bestimmen sie Erwartungswert, Varianz und Standardabweichung für die Augenzahl bei einem Wurf.

10. Eine Stadt ist im 20. Jahrhundert einmal überschwemmt worden. Wie viele Überschwemmungen pro Jahrhundert muss ein Klimamodell mindestens vorhersagen, damit es durch die Beobachtung nicht mit 95%iger Wahrscheinlichkeit **ausgeschlossen** werden kann?

11. Welches 90% Vertrauensintervall $\mu_1,...\mu_2$ gehört unter der Annahme einer nach der Poisson verteilten Anzahl zu einem einzigen beobachteten Ereignis?

Aufgaben zu kontinuierlichen Verteilungen

12. Bitte zeigen Sie, dass die Varianz einer gleich-verteilten Größe $V=(x_{max}-x_{min})^2/12$ beträgt!

13. Radioaktive Elemente zerfallen gemäß $N=N_0 e^{-t/\tau}$, wobei N_0 die Anzahle der Atome zum Zeitpunkt $t=0$ ist. Wie lautet die dazu gehörige Verteilungsfunktion?

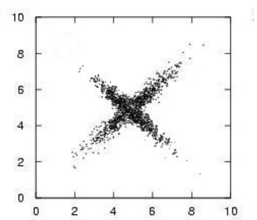
14. Ein perfekt linearer 10 Bit Analog-Digitalwandler (ADC) für den Bereich 0 bis 5 Volt ist so geeicht, dass der Ausgang (0000000000) „mindestens 0 Volt“ und (1111111111) „höchstens 5 Volt“ bedeuten. Welcher Spannung und welcher Standard-Abweichung entspricht der Messwert (1000000000)?

15. Die so genannte Fehlerfunktion („Error Function“) ist durch $erf(x)=\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$ gegeben. Was ist $erf(-x)$? In welchem Zusammenhang steht diese Funktion zur kumulativen Standard-Normalverteilung $\Phi(z)$?

Aufgaben zu Verteilungen in mehreren Dimaensionen

16. Bitte zeigen Sie, dass für unkorrelierte Variablen die folgende Gleichung gilt: $E(x\cdot y)=\iint x\cdot y\cdot p(x,y)dx\,dy=\mu_x\cdot\mu_y$!

17. Das folgende Bild zeigt die *(x,y)* Wertepaare einer Messreihe: Sind die Daten korreliert? Bitte schätzen sie den Korrelationskoeffizienten ab!



18. Die Dicke eines aus zwei miteinander reagierenden Materialien bestehenden Verbundmaterials wird gemessen, indem die beiden Lagen einzeln vermessen werden. In einer 100 Mess-Paare langen Messreihe wird festgestellt, dass die jeweiligen Einzelmessungen zu 50 % anti-korreliert sind. Wenn die Streubreite der Einzelmessungen jeweils $\sigma_{x1}=\sigma_{x2}=10\,\mu m$ beträgt, mit welcher Genauigkeit ist dann die durchschnittliche Gesamtdicke des Verbundmaterials bekannt?

19. Es seien a,b,c und d voneinander unabhängige Messgrößen. Finden Sie einen möglichst eleganten Ausdruck für die die Varianz der Funktion $y=\frac{a\cdot b}{c\cdot d}$!

20. Ein Professor der Fachhochschule Münster versucht die durchschnittliche Länge eines Tages zu messen, indem er erst die Dauer des Tages misst, dann die Dauer der Nacht und schließlich beides addiert. Dabei ist die größte statistische Unsicherheit Wetter bedingt: bei gutem Wetter überschätzt er die Tageslänge, bei schlechtem Wetter die Länge der Nacht. Welchen Korellationskoeffizienten erwarten Sie für die beiden Messgrößen?

Antworten:

1. Ein Bernoulli Experiment ist ein Zufalls-Experiment, bei dem nur zwei Resultate (gut, schlecht), (ja, nein),...möglich sind. Ist die Wahrscheinlichkeit, dass das erste Resultat eintritt *p* und wird das Experiment *n* mal wiederholt, so erhält man die Binomialverteilung *P (n , k)* .

2. Laut Bernoulli ist $P(n,k)=\binom{n}{k}\cdot p^k(1-p)^{(n-k)}$, also
   $P(10{,}6)=\binom{10}{6}\cdot 0{,}5^{10}=0{,}20508$, $P(10{,}7)=\binom{10}{7}\cdot 0{,}5^{10}=0{,}1172$,
   $P(10{,}8)=0{,}0439$, $P(10{,}9)=0{,}0098$, $P(10{,}10)=0{,}5^{10}=0{,}001$. Die Summe dieser Wahrscheinlichkeiten ist *38,6%.*

3. Es gibt drei Zweierkombinationen und drei verschiedene Einzelkreuzchen, zusammen also 6 Möglichkeiten, von denen nur eine richtig ist: $p=1/6$. Es sind ferner *n = 8* und $k=8,7,6,$ oder *5*. für die Noten *1,2,3* oder *4*. Daher ist die Wahrscheinlichkeit
   für eine „1“ $P(8{,}8)=\binom{8}{8}\cdot(1/6)^8(5/6)^0=1\cdot(1/6^8)\cdot 1\approx 5{,}95\cdot 10^{-7}$,
   für eine „2“ $P(8{,}7)=\binom{8}{7}\cdot(1/6)^7(5/6)^1=8\cdot(5/6^8)\approx 2{,}38\cdot 10^{-5}$,
   für eine „3“ $P(8{,}6)=\binom{8}{6}\cdot(1/6)^6(5/6)^2=28\cdot(5^2/6^8)\approx 4{,}17\cdot 10^{-4}$ und
   für eine „4“ $P(8{,}5)=\binom{8}{5}\cdot(1/6)^5(5/6)^3=56\cdot(5^3/6^8)\approx 4{,}16\cdot 10^{-3}$.
   Insgesamt ist die Chance durchzufallen 99,53 %...

4. Es ist $\binom{n}{n}=\frac{n!}{0!n!}=\binom{n}{0}=1$, $\binom{n}{k}=\frac{n!}{k!(n-k)!}=\binom{n}{(n-k)}$ und $\binom{n}{1}=n$.

5. Hier hilft die geometrische Verteilung: $P=(1-p)\cdot(1-p)\cdot\ldots\cdot p=(1-p)^{(k-1)}\cdot p$, also für *k* = *1* ist $P=(1-p)^0\cdot p=4\,\%$ und für *k* = *25* ist $P=(0{,}96)^{24}\cdot 0{,}04=1{,}5\,\%$.

6. Dies kann nur heißen: 0 gefundene Teilchen haben nach Poisson Statistik nur eine Wahrscheinlichkeit von 5%, oder $P(\mu,k)=\frac{\mu^k}{k!}e^{-\mu}$, also $5\%=P(\mu,0)=e^{-\mu}$. Aufgelöst nach μ ergibt dies $\mu\approx 3$. Man müsste also 30 Millionen Reaktionen untersuchen und kein einziges Teilchen finden.

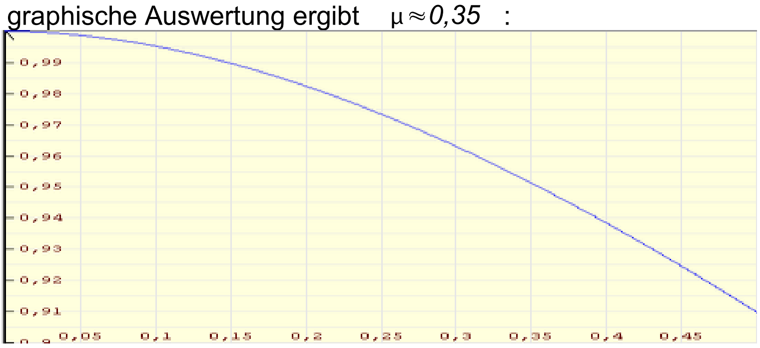
7. Für die erste Frage hilft die negative Binomialverteilung, denn man kann umformulieren: jeder Tag ist ein Versuch, und ein Regentag ist ein günstiges Ereignis... $P(30{,}24)=\binom{29}{23}\cdot 0{,}8^{24}(0{,}2)^{(6)}=475020\cdot 0{,}0047\cdot 0{,}00064=0{,}1435$.
   Kurz: *k* ist fest und *n* gesucht.
   Die zweite Frage zielt auf die „normale“ Binomialverteilung ab:
   $P(30{,}24)=\binom{30}{24}\cdot 0{,}8^{24}(0{,}2)^{(6)}=593775\cdot 0{,}0047\cdot 0{,}00064=0{,}1795$

8. Hier hilft die geometrische Verteilung: Es ist $E(k)=1/p=6$ und $V(k)=(1-p)/p^2=(5/6)/(1/36)=30$, $\sigma_k=\sqrt{30}$. ???? Genau zwei weiße Weihnachten heißt; davor nicht, dann zwei mal weiß, dann wieder nicht: $P=(1-p)\cdot p\cdot p\cdot(1-p)=(1/36)\cdot(25/36)=1{,}93\,\%$. Es wir im Jahrtausend nur 19 mal passieren.???? Normierung?

9. Mittelwert: $E(k)=(1+2+3+4+5+6)/6=21/6=3{,}5$.

Varianz $V(k)=\frac{2}{6}(0{,}5^2+1{,}5^2+2{,}5^2)=2{,}916$, Standardabweichung $\sigma_k=1{,}71$

10. Es muss gelten $P(\mu,0)+P(\mu,1)\geq 0{,}95$, also $\frac{\mu^0}{0!}e^{-\mu}+\frac{\mu^1}{1!}e^{-\mu}\geq 0{,}95$. Die näherungsweise Lösung dieser transzendenten Gleichung folgt über eine Reihenentwicklung der Exponentialfunktion: $e^{-\mu}\approx 1-\mu+\mu^2/2$ und liefert $\mu\approx\sqrt{1/10}=0{,}32$ Ein Modell, welches weniger als eine Überschwemmung pro 300 Jahre voraussagt ist mit 95%iger Wahrscheinlichkeit ausgeschlossen.
Eine graphische Auswertung ergibt $\mu\approx 0{,}35$:



11. Es gilt, die 95% Untergrenze zu finden - und die 95% Obergrenze.
Untergrenze: $P(\mu_1,0)=\frac{\mu_1^0}{0!}e^{-\mu_1}=0{,}95\Rightarrow\mu_1=0{,}051$,
Obergrenze $P(\mu_2,0)+P(\mu_2,1)=\frac{\mu_2^0}{0!}e^{-\mu_2}+\frac{\mu_2^1}{1!}e^{-\mu_2}=(1+\mu_2)e^{-\mu_2}=0{,}05\Rightarrow\mu_2=4{,}74$
(Die Obergrenze wurde numerisch mit der Seite http://www.arndt-bruenner.de/mathe/java/plotter.htm bestimmt.) Diese beiden Grenzen markieren das 90 % Konfidenzintervall.

12. Zur Berechnung von $V=\int p(x)\cdot(x-\mu)^2dx$ wählen wir den Spezialfall $\mu=0$ und berechnen $V=\frac{1}{2a}\int_{-a}^{a}x^2dx=\frac{1}{3}a^2$, wobei a die halbe Intervallbreite ist: $a=(x_{max}-x_{min})/2$. Einsetzen ergibt $V=\frac{1}{12}(x_{max}-x_{min})^2$.

13. Um $p(x)=\frac{1}{n}\cdot\frac{dn}{dx}$ zu erhalten differenzieren wir $\frac{d\,e^{-t/\tau}}{dt}0=-\frac{1}{\tau}\cdot e^{-t/\tau}$ und normieren anschließend auf 1. Ergebnis: $p(t)=\frac{1}{\tau}\cdot e^{-t/\tau}$.

14. Jede Änderung des Letzten Bits entspricht einem Sprung von $\Delta U=5\,Volt/(2^{10})\approx 5\,mV$. Der Sprung (0111...) auf (100...) ist (wie man sich mit 1 und 2 Bits leicht veranschaulichen kann) genau die Mitte des Messbereiches. Dazwischen muss eine Gleichverteilung angenommen werden. Daher ist $U_{gemessen}=2{,}5025\,V$ und $\sigma_U=5mV/\sqrt{12}\approx 1{,}44\,mV$

15. Da $erf(x)=\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$ bleibt beim Vorzeichenwechsel von x der Integrand positiv und vom Betrag her unverändert. Das Integral wird aber von 0 abwärts gerechnet, so dass insgesamt ein Minuszeichen übrig bleibt: $erf(-x)=-erf(x)$.

Ein Vergleich mit $\Phi(z)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}e^{-t^2/2}dt$ erfolgt in zwei Schritten. Zunächst wird der negative Teil gesondert gerechnet, denn sein Integral ist immer ½: $\Phi(z)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0}e^{-t^2/2}dt+\frac{1}{\sqrt{2\pi}}\int_0^z e^{-t^2/2}dt=\frac{1}{2}+\frac{1}{\sqrt{2\pi}}\int_0^z e^{-t^2/2}dt$ Dann wird der Integrand „gleich geschaltet“ : Setze $t^2/2=x^2$. Dann wird die obere Grenze

$z/\sqrt{2}=x_o$ und das Differenzial $dt/\sqrt{2}=dx$. Daher ist $\Phi(z)=\frac{1}{2}+\frac{1}{\sqrt{2\pi}}\int_0^z e^{-t^2/2}dt=\frac{1}{2}+\frac{1}{\sqrt{2\pi}}\int_0^{z/\sqrt{2}} e^{-x^2}(\sqrt{2}dx)$. Zieht man die Wurzel nach vorne, so wird daraus $\Phi(z)=\frac{1}{2}+\frac{1}{\sqrt{\pi}}\int_0^{z/\sqrt{2}} e^{-x^2}dx$. Ein Vergleich mit $erf(x)$ gibt das Gesamtergebnis: $\Phi(z)=\frac{1}{2}+\frac{1}{2}\cdot erf(\frac{z}{\sqrt{2}})$.

16. Wenn die Variable unkorreliert sind, dann können wir schreiben: $E(x\cdot y)=\iint x\cdot y\cdot p(x,y)dx\,dy=\mu_x\cdot\mu_y=\int x\,p_x(x)dx\cdot\int y\,p_y(y)dy=\mu_x\cdot\mu_y$

17. Die Struktur macht deutlich: diese Daten sind stark korreliert! Allerdings werden sich die Beiträge zur Kovarianz in etwa zu 0 summieren – ein Beispiel dafür, dass die Kovarianz auch bei korrelierten Daten 0 sein kann.

18. Es gilt $y=x_1+x_2\rightarrow\sigma_y^2=\sigma_{X1}^2+2\,cov(x_{1,}x_2)+\sigma_{X2}^2$. Da die Standardabweichungen gleich groß sind, kann man vereinfachen: $\sigma_y^2=2\sigma_X^2+2\sigma_X^2\cdot\rho_{x1,x2}$, also hier $\sigma_y^2=2\cdot(10\,\mu m)^2+2\cdot(10\,\mu\,m)^2\cdot(-1/2)=(10\,\mu\,m)^2$. Der Erwartungswert für y ist um den Faktor $\sqrt{n}$ genauer bekannt. Also kennt man die durchschnittliche Gesamtdicke auf $\sigma_{\mu_y}=\sigma_X/\sqrt{100}=1\,\mu\,m$ genau.

19. Da die Größen unkorreliert sind, können wir $\sigma_y^2=\sum_{i=1}^{n}\left[\frac{\partial y}{\partial x_i}\right]_{x_i=\mu_i}^2\cdot\sigma_i^2$ benutzen. Dabei ist $\frac{\partial y}{\partial a}=\frac{b}{c\cdot d}=\frac{y}{a}$ und entsprechend für b. Die Terme für den Nenner sind $\frac{\partial y}{\partial c}=-\frac{a\cdot b}{c^2\cdot d}=\frac{-y}{c}$ und entsprechend für d. Beim Quadrieren verschwinden die Minuszeichen, und es bleibt $\left(\frac{\sigma_y}{y}\right)^2=\left(\frac{\sigma_a}{a}\right)^2+\left(\frac{\sigma_b}{b}\right)^2+\left(\frac{\sigma_c}{c}\right)^2+\left(\frac{\sigma_d}{d}\right)^2$, die „Quadratische Addition der relativen Fehler“

20. Da die Summe immer 24 h beträgt, sind beide Messungen zu 100% anti-korreliert.