UNIVERSITY OF OSLO

Faculty of Mathematics and Natural Sciences

Exam in GEO4310 – Stochastic methods in hydrology
Day of exam: 8 December 2015
Exam hours: 09:00 – 12:00
This examination paper consists of 5 pages.
Appendices: No
Permitted materials: Calculator

Make sure that your copy of this examination paper is complete before answering.

Task 1 Definitions (10 points):

Give a brief definition for each of the following terms:

a) Coefficient of variation
b) Central limit theorem
c) Autocorrelation
d) Confidence interval
e) Homoscedasticity

Answers

a) Coefficient of variation is the ratio of the sample standard deviation to the sample mean:
   Cv = s/x (page 29 in the compendium from 2013)
b) Central limit theorem: If Xn are independent, identically distributed (need not to be normal) random variables having mean μ and finite nonzero variance σ^2, a sum of X's is normally distributed with mean n·μ and variance n·σ^2. (p 50)
c) Autocorrelation is the correlation between two copies of the same time-series but with a time shift of L: ρL = cor (xt, xt + L) (page 130 in the compendium from 2013)
d) Confidence interval is a range of values that has a specified probability of containing the parameter being estimated. This statement may be written as P(L<θ<U) = 1-α
e) The variance of a random variate is constant across a range of values.

Task 2 Distributions and exreme value statistics (~30 points)

a) *Two catchments receive the same annual precipitation, 1000 mm/year. Based on 30 years of streamflow data, we have estimated the following statistics:*

	Catchment 1	*Catchment 2*
Mean (m^3/s)	*0.4*	*0.6*
Standard deviation (m^3/s)	*0.2*	*0.1*
Skewness	*3.5*	*0.0*

Sketch the probability distributions (pdf) of catchment 1 and 2 in the same figure. Sketch the cumulative probabilitiy distributions (cdf) of the two catchments in a different figure. Which mathematical distributions could fit catchment 1 and which could fit catchment 2?

b) *The probability that an annual maximum flood at a measuring site in a river exceeds 100 m^3/s is $p=0.1$, and the probability of not exceeding 100 m^3/s is $q=1-p=0.9$. Assume that annual maximum floods are independent of each other.*

What is the probability that the annual maximum flood exceeds 100 m^3/s excactly three times during a period of 30 years?

c) *Flood-frequency analysis is used to calculate the magnitude of a flood.*
   i) *Describe and sketch how you may select data used for a flood-frequency analysis from a data series consisting of daily mean flow.*

   ii) *How is the return period defined? What is the relationship between the return period (T) and the cumulative distribution function (F), and which non-exceedence probability does a 100-year flood have?*

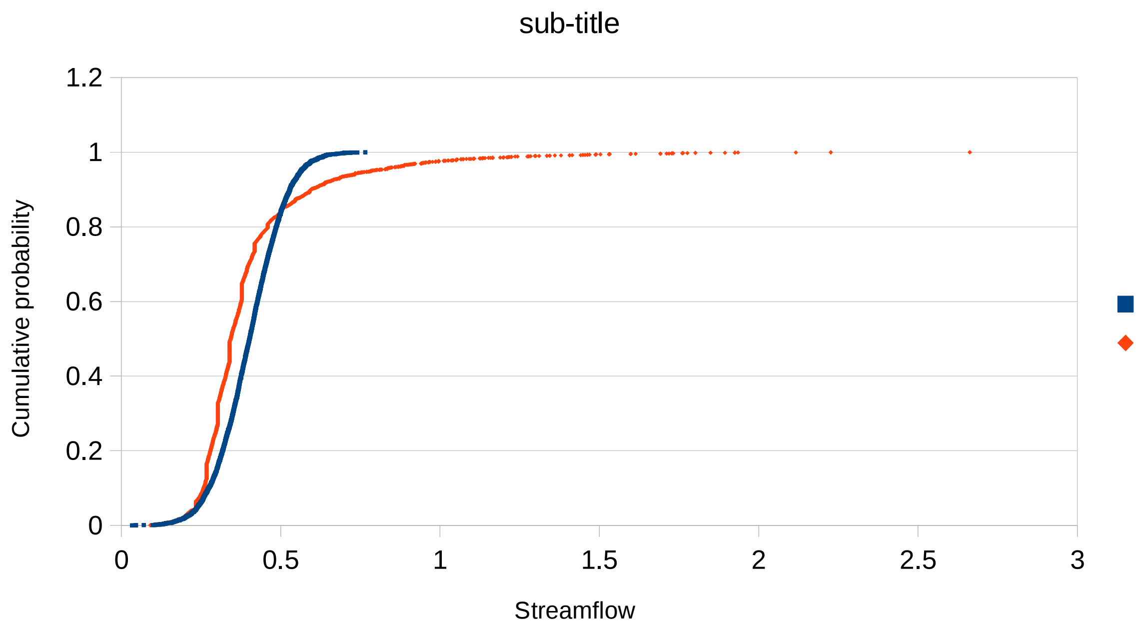
   iii) *For a certain catchment, extreme streamflow values (q [m^3/s]) may be described by a Gumbel distribution with parameters $\alpha = 0.234$ og $\beta = 0.369$.*

$F(q) = \exp(-\exp(-(q-\beta)/\alpha))$

How large is a 100-year flood in this catchment?

Answer to a)

start med å markere gjennomsnittsverdien (middelverdien) for hvert felt på x-aksen.
Så kan du tegne inn topp-punktet på kurven. Hvis fordelingen er skjev ligger denne verdien litt til høyre for topp-punktet. Hvis fordelingen er symmetrisk (skjevhet = 0)
ligger topp-punktet rett over middelverdien. Jo større standardavvik, jo bredere og lavere kurve.



Answer to b)

Use the binomial coefficient to find the number of combinations that the annual maximum flood can exceed excactly three times during a period of 30 years:

C n,k =

n!

/ [(n − k)!k!]

C 30,3 =

30!

/ [(30 − 3)!3!] = 30!

*/ [27! 3!] = 28*29*30 / (1*2*3) = 4060*

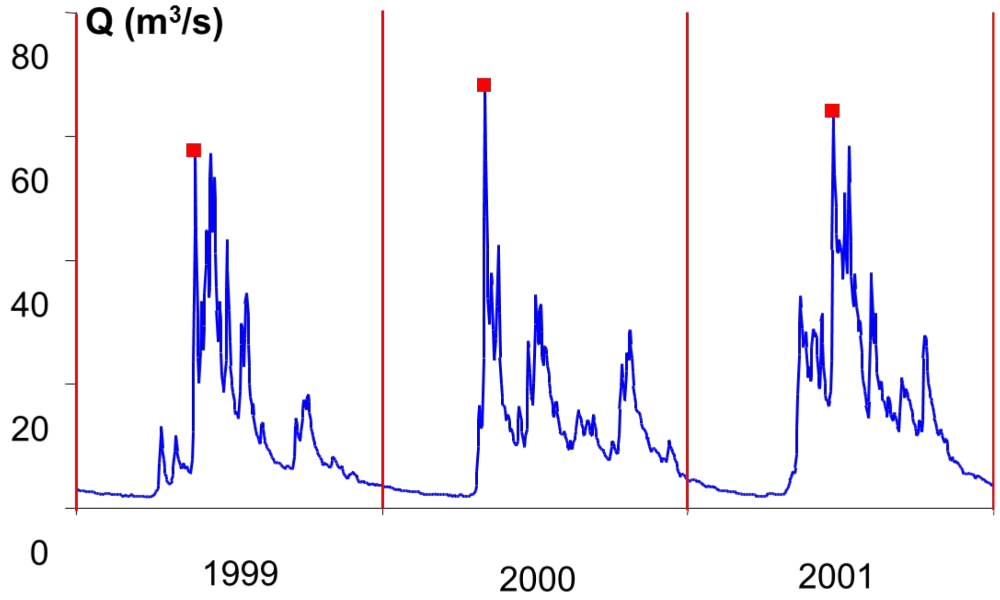
(Here, the students need to abbreviate the expression before they insert the factorials into their calculators)

To get the probability, t*his number is inserted into the formula* $C\ 30,3*p^k(1-p)^{n-k}$*, which needs to be reasoned in the following way:*
For each realization, there are 3 successes among 30 outcomes: $p^3(1-p)^{27}$

Multiplying in the total number of combinations, we get the probability that we get exactly three ten-year floods in 30 years:

$C\ 30,3*p^k(1-p)^{n-k}$
$= 4060*p^3(1-p)^{27}$
$= 4060*0.1^3 0.9^{27}$
$= 0.24$
The probability that we get exactly three ten-year floods in 30 years is 0.24.

Answer to c i)
Annual maximum series (AMS): Velger ut den største verdien hvert år. Verdiene må være uavhengige, så hvis én ekstrem flomhendelse opptrer både 31. desember (f.eks.1990) og 1. jan (f.eks.1991), tilfaller denne flomhendelsen det året med høyest døgnmiddelverdi. I det andre året må en annen flomhendelse velges (nevnt på gruppetime i mai 2013).
Ulempe med AMS: hvis det opptrer flere ekstreme hendelser i løpet av ett år, blir bare den største verdien tatt med i dataserien. I andre år hender det at det ikke opptrer ekstreme flommer i det hele tatt, men den høyeste (ikke ekstreme) vannføringen tas likevel med i dataserien.



Answer to c ii)

Gjentaksintervall for en gitt vannføring er gjennomsnittlig antall år mellom hver gang den årlige maksimale døgnmidddelvannføringen overskrider den aktuelle vannføringen. Gjentaksintervallet er den inverse av overskridelsessannsynligheten:

$T = 1/E = 1/(1-F)$

En 100-årsflom har gjentaksintervall 100 år. Vi løser ut F og setter inn i likningen:

$1-F = 1/T \;\; \rightarrow F = 1-(1/T) = 1-0.01 = 0.99.$

Totalt 4 poeng, 1 for riktig beskrivelse, 1 for likning, 1 for riktig F, og 1 for å regne ut F.

Answer to c iii)

Vi må snu ligningen for å uttrykke Q som funksjon av F:

$F(q) = exp(-exp(-(q-\beta)/\alpha))$

$-exp(-(q-\beta)/\alpha) = ln(F)$

$-(q-\beta)/\alpha = ln(- ln(F))$

$(q-\beta) = - \alpha \, ln(- ln(F))$

$q = \beta - \alpha \, ln(- ln(F))$

Setter inn F = 0.99, samt $\alpha = 0.234$ *og* $\beta = 0.369$ *og får* $q = 1,45 \, m^3/s$

Task 3 Hypothesis testing (~30 points)

a) Is there a significance difference between (a) the annual mean precipitation and (b) the variance of precipitation in point A and point B at a level of significance of 5% if for a 50-year period it has been found that:

Point	**Precipitation** Mean	Standard deviation
A	720 mm	220 mm
B	660 mm	140 mm

b) Based on 90-year records of yearly maximum discharge data at two adjacent stations, the following statistics are calculated (with n=90). The mean value in station A is 130 m3/s, standard deviation in A is 50 m3/s. The mean value in station B is 120 m3/s, standard deviation in B is 50 m3/s. The significance level is 5%

	Q≤50	50<Q≤100	100<Q≤150	150<Q≤250	Q>250
Station A (# years)	6	24	35	15	10
Station B (# years)	5	25	20	26	14

Use Chi square method to test:

i) the hypothesis that data in Station A and Station B have the same distribution

ii) the hypothesis that data in Station A has normal distribution.

Solution Q2

a) $\overline{X}_A = 720$ $\overline{X}_B = 660$

$S_A = 220$ $S_B = 140$

Ho: $\mu_1 = \mu_2$ Ha: $\mu_1 \neq \mu_2$

$$T' = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{720 - 660}{\sqrt{\frac{220^2}{50} + \frac{140^2}{50}}} = 1.627$$

Ho rejects if T' > $(w_A t_A + w_B t_B)/(w_A + w_B)$

Where $w_A = S_A^2 / n_A = 48400/50 = 968$

$w_B = S_B^2 / n_B = 19600/50 = 392$

for $\alpha = 5\%$

$t_A = t_{1-\alpha/2, n_A - 1} = t_{0.975, 49} = 2.01 = t_B$ (shouldn't it be 2.3124?) 2.01 is for 95% only

and $\dfrac{w_A t_A + w_B t_B}{w_A + w_B} = \ldots = 2.01 > T' = 1.627$

Therefore Ho cannot be rejected, i.e. both gauges receive the same amount of rain at the significance level of 5%.

b) F-test

The F statistic is a simple ratio of the two unbiased estimator of the variance:

$$F_c = s_1^2 / s_2^2$$

The hypothesis of having the same variace will be rejected if

$F_c > F_{1-\alpha, n1-1, n2-1}$

$F_c = s_1^2 / s_2^2 = \dfrac{48400}{19600} = 2.47$

$F_{1-\alpha, n1-1, n2-1} = F_{0.95, 49, 49} \approx 1.7$ (Any value between 1.74 to 1.45 is correct, since it is interpolated from Table)

Since $F_c > F_{1-\alpha, n1-1, n2-1}$

So the difference between the variances of two stations is significant!

Solution Q3:

(A)

Q≤50	50<Q≤100	100<Q≤150	150<Q≤250	Q>250		
6	24	35	15	10	Oj	
5	25	20	26	14	Ej	90
1	-1	15	-11	-4	(Qj-Ej)	
1	1	225	121	16	(Oj-Ej)^2	364
0.2	0.04	11.25	4.653846	1.1428 57	(Oj- Ej)^2/Ej	X^2_c=17.2867

$X^2_{0.05,5-1}$ =9.488 < X^2_c=17.28

Decision: the hypothesis of A and B belonging to same distribution is rejected.

(B)

Q≤50	50<Q≤100	100<Q≤150	150<Q≤250	Q>250		
6	24	35	15	10		Oj
0.0548	0.2195	0.3881	0.3364	0.0082	1.007	Expected rel freq
4.932	19.755	34.929	30.276	0.738		Ej
0.23127	0.912175	0.000144	7.707629	116.239 4	(Oj-Ej)^2/Ej	X^2_c =125.090 6

P(Q<50) = p(z<(50-130)/50)=p(z<-1.6) = 1-0.9452=0.0548

p(Q<100) = p(z<(100-130)/50=p(z<-0.6)=1-0.7257=0.2743
p(50<Q<100)=0.2743-0.0548 =
0.2195
p(Q<150) = p(z<(150-
130)/50=p(z<0.4)=0.6554
p(100<Q<150) = 0.6554-0.2743 =
0.3811
p(Q<250) = p(z<(250-
130)/50)=p(z<2.4)=0.9918
p(150<Q<250) = 0.9918-0.6554 =
0.3364
p(Q>250) =1-P(Q<250) = 1-
0.9918=0.0082

$X^2_{0.05,5-2-1}$ =5.991 < X^2_c=125.09

Decision: The hypothesis of A is normally distributed is rejected.

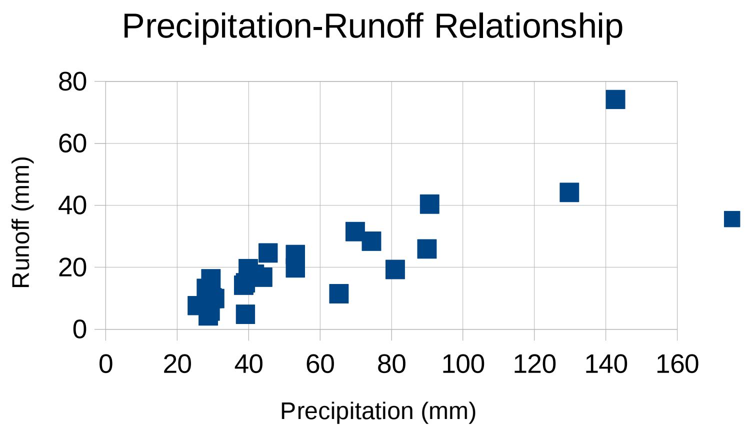
Task 4 Regression and time series (~30 points)

a) *The precipitation and runoff data for 25 Storms on the Monocacy River are provided in the table below.*
   i) *Plot the observed data for runoff vs. Precipitation*
   ii) *Determine and draw the regression line of runoff on precipitation (that is, the mean runoff for given value of precipitation).*
   iii) *Estimate the variance of runoff for a given precipitation. Assume that the variance of th runoff is constant with precipitation.*
   iv) *Assume that the runoff corresponding to a given precipitation is a normal variate; what is the probability that the runoff will exceed 50 mm. During a storm with 100-mm. precipitation?*

Storm ID	Precipitation (mm.)	Runoff (mm.)
1	28.194	13.208
2	29.718	10.16
3	45.466	24.638
4	142.748	74.168
5	28.702	4.318
6	39.116	4.826
7	81.026	19.304
8	43.942	16.764
9	53.086	19.812
10	69.85	31.496
11	30.48	9.906
12	25.654	7.62
13	41.656	17.78
14	39.878	19.558
15	39.116	14.986
16	53.086	24.13
17	89.916	25.908
18	29.718	9.906
19	29.21	5.842
20	65.278	11.43
21	90.678	40.386
22	129.794	44.196
23	38.608	14.224
24	74.422	28.448
25	29.464	16.256

Answers

i)



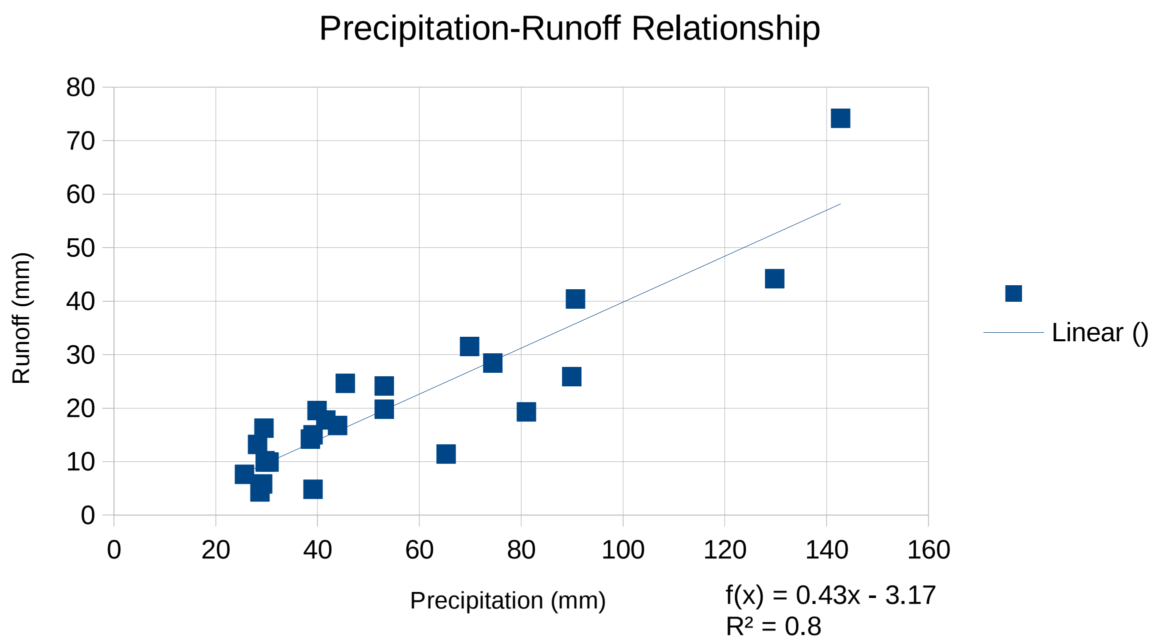
ii)

x_avg = 1368.81/25 = 54.75

y_avg = 509.3/25 = 20.37

Beta = \sum(xiyi) – n x_avg y_avg / \sum(xi^2) – n x_avg^2 = (38216.7-25(54.75)(20.37)) / 98978.96 – 25(54.75)^2 = 0.43

alpha = y_avg – Beta x_avg = 20.37 – (0.43)(54.75) = -3.1725



iii)

S2Y|x = 1 / n-2 * \sum(y_i – y*_i)^2 = 1104.92 / 25-2 = 42.2 mm^2

SY|x = \sqrt(S2Y|x) = 6.5 mm.

iv) When the precipitation is 100 mm., the mean runoff:

$E(Y \mid X=100) = -3.17 + 0.43(100) = 39.83$ mm.

The normal distribution for the runoff (Y) in this storm is N(39.8, 6.5)

And $P(Y>50 \mid X=100) = 1 - \phi(50 – 39.83) = 1 - \phi(20.97/6.5) = 1 - 0.94 = .06$