Concentration Inequalities

Emin Karayel and Yong Kiam Tan*

December 6, 2023

Abstract

Concentration inequalities provide bounds on how a random variable (or a sum/composition of random variables) deviate from their expectation, usually based on moments/independence of the variables.

The most important concentration inequalities (the Markov, Chebyshev, and Hoelder inequalities and the Chernoff bounds) are already part of HOL-Probability. This entry collects more advanced results, such as Bennett's/Bernstein's Inequality, Bienaymé's Identity, Cantelli's Inequality, the Efron-Stein Inequality, McDiarmid's Inequality, and the Paley-Zygmund Inequality.

Contents

1 Preliminary results

theory *Concentration-Inequalities-Preliminary*
  imports *Lp.Lp*
begin

Version of Cauchy-Schwartz for the Lebesgue integral:

*The authors contributed equally to this work.

lemma *cauchy-schwartz*:
  fixes $f\ g$ :: - $\Rightarrow$ *real*
  assumes $f \in$ *borel-measurable* M $g \in$ *borel-measurable* M
  assumes *integrable* M $(\lambda x.\ (f\ x)\ \hat{}2)$ *integrable* M $(\lambda x.\ (g\ x)\ \hat{}2)$
  shows *integrable* M $(\lambda x.\ f\ x * g\ x)$ (**is** *?A*)
    $(\int x.\ f\ x * g\ x\ \partial M) \leq (\int x.\ (f\ x)\hat{}2\ \partial M)$ *powr* $(1/2) * (\int x.\ (g\ x)\hat{}\ 2\ \partial M)$ *powr* $(1/2)$
    (**is** $?L \leq ?R$)
⟨*proof*⟩

Generalization of *prob-space.indep-vars-iff-distr-eq-PiM′*:

lemma (**in** *prob-space*) *indep-vars-iff-distr-eq-PiM′′*:
  fixes I :: *′i set* **and** X :: $'i \Rightarrow 'a \Rightarrow 'b$
  assumes *rv*: $\bigwedge i.\ i \in I \Longrightarrow$ *random-variable* $(M'\ i)\ (X\ i)$
  shows *indep-vars* $M'\ X\ I \longleftrightarrow$
    distr M $(\Pi_M\ i{\in}I.\ M'\ i)$ $(\lambda x.\ \lambda i{\in}I.\ X\ i\ x) = (\Pi_M\ i{\in}I.$ *distr* $M\ (M'\ i)\ (X\ i))$
⟨*proof*⟩

lemma *proj-indep*:
  assumes $\bigwedge i.\ i \in I \Longrightarrow$ *prob-space* $(M\ i)$
  shows *prob-space.indep-vars* (*PiM* $I\ M$) M $(\lambda i\ \omega.\ \omega\ i)$ I
⟨*proof*⟩

lemma *forall-Pi-to-PiE*:
  assumes $\bigwedge x.\ P\ x = P$ (*restrict* $x\ I$)
  shows $(\forall x \in Pi\ I\ A.\ P\ x) = (\forall x \in PiE\ I\ A.\ P\ x)$
  ⟨*proof*⟩

lemma *PiE-reindex*:
  assumes *inj-on* $f\ I$
  shows *PiE* I $(A \circ f) = (\lambda a.$ *restrict* $(a \circ f)\ I)$ ‘ *PiE* $(f$ ‘ $I)$ A (**is** *?lhs* = *?g* ‘ *?rhs*)
⟨*proof*⟩

context *prob-space*
begin

lemma *indep-sets-reindex*:
  assumes *inj-on* $f\ I$
  shows *indep-sets* A $(f$ ‘ $I)$ = *indep-sets* $(\lambda i.\ A\ (f\ i))$ I
⟨*proof*⟩

lemma *indep-vars-reindex*:
  assumes *inj-on* $f\ I$
  assumes *indep-vars* $M'\ X'$ $(f$ ‘ $I)$
  shows *indep-vars* $(M' \circ f)$ $(\lambda k\ \omega.\ X'\ (f\ k)\ \omega)$ I
  ⟨*proof*⟩

lemma *indep-vars-cong-AE*:
  assumes $AE\ x\ in\ M.\ (\forall\, i \in I.\ X'\ i\ x = Y'\ i\ x)$
  assumes *indep-vars* $M'\ X'\ I$
  assumes $\bigwedge i.\ i \in I \Longrightarrow$ *random-variable* $(M'\ i)\ (Y'\ i)$
  shows *indep-vars* $M'\ Y'\ I$
⟨*proof*⟩

end

Integrability of bounded functions on finite measure spaces:

lemma *bounded-const*: *bounded* $((\lambda x.\ (c{::}real))\ \text{‘}\ T)$
  ⟨*proof*⟩

lemma *bounded-exp*:
  fixes $f :: 'a \Rightarrow real$
  assumes *bounded* $((\lambda x.\ f\ x)\ \text{‘}\ T)$
  shows *bounded* $((\lambda x.\ exp\ (f\ x))\ \text{‘}\ T)$
⟨*proof*⟩

lemma *bounded-mult-comp*:
  fixes $f :: 'a \Rightarrow real$
  assumes *bounded* $(f\ \text{‘}\ T)$ *bounded* $(g\ \text{‘}\ T)$
  shows *bounded* $((\lambda x.\ (f\ x) * (g\ x))\ \text{‘}\ T)$
⟨*proof*⟩

lemma *bounded-sum*:
  fixes $f :: 'i \Rightarrow 'a \Rightarrow real$
  assumes *finite* I
  assumes $\bigwedge i.\ i \in I \Longrightarrow$ *bounded* $(f\ i\ \text{‘}\ T)$
  shows *bounded* $((\lambda x.\ (\sum i \in I.\ f\ i\ x))\ \text{‘}\ T)$
  ⟨*proof*⟩

lemma (**in** *finite-measure*) *bounded-int*:
  fixes $f :: 'i \Rightarrow 'a \Rightarrow real$
  assumes *bounded* $((\lambda\ x.\ f\ (fst\ x)\ (snd\ x))\ \text{‘}\ (T \times space\ M))$
  shows *bounded* $((\lambda x.\ (\int \omega.\ (f\ x\ \omega)\ \partial M))\ \text{‘}\ T)$
⟨*proof*⟩

lemmas *bounded-intros* =
  bounded-minus-comp bounded-plus-comp bounded-mult-comp bounded-sum finite-measure.bounded-int
  bounded-const bounded-exp

lemma (**in** *prob-space*) *integrable-bounded*:
  fixes $f :: \text{-} \Rightarrow ('b :: \{banach, second\text{-}countable\text{-}topology\})$
  assumes *bounded* $(f\ \text{‘}\ space\ M)$
  assumes $f \in M \rightarrow_M borel$
  shows *integrable* $M\ f$
⟨*proof*⟩

end

2 Bennett's Inequality

In this section we verify Bennett's inequality [1] and a (weak) version of Bernstein's inequality as a corollary. Both inequalities give concentration bounds for sums of independent random variables. The statement and proofs follow a summary paper by Boucheron et al. [2].

theory *Bennett-Inequality*
  imports *Concentration-Inequalities-Preliminary*
begin

context *prob-space*
begin

lemma *indep-vars-Chernoff-ineq-ge*:
  assumes I: *finite* I
  assumes *ind*: *indep-vars* (λ -. *borel*) X I
  assumes *sge*: $s \geq 0$
  assumes *int*: $\bigwedge i.\ i \in I \Longrightarrow$ *integrable* M ($\lambda x.$ *exp* ($s * X\ i\ x$))
  shows *prob* $\{x \in$ *space* $M.\ (\sum i \in I.\ X\ i\ x\ -$ *expectation* $(X\ i)) \geq t\} \leq$
    exp $(-s{*}t)$ $*$
    $(\prod i{\in}I.$ *expectation* $(\lambda x.$ *exp*$(s * (X\ i\ x\ -$ *expectation* $(X\ i)))))$
⟨*proof*⟩

definition *bennett-h*::*real* ⇒ *real*
  where *bennett-h* $u = (1 + u) *$ *ln* $(1 + u) - u$

lemma *exp-sub-two-terms-eq*:
  fixes x :: *real*
  shows *exp* $x - x - 1 = (\sum n.\ x\hat{}(n{+}2)$ / *fact* $(n{+}2))$
    summable $(\lambda n.\ x\hat{}(n{+}2)$ / *fact* $(n{+}2))$
⟨*proof*⟩

lemma *psi-mono*:
  defines $f \equiv (\lambda x.$ (*exp* $x - x - 1) - x\hat{}2$ / $2)$
  assumes *xy*: $a \leq (b$::*real*)
  shows $f\ a \leq f\ b$
⟨*proof*⟩

lemma *psi-inequality*:
  assumes *le*: $x \leq (y$::*real*) $y \geq 0$
  shows $y\hat{}2 *$ (*exp* $x - x - 1) \leq x\hat{}2 *$ (*exp* $y - y - 1)$

⟨*proof*⟩

lemma *bennett-inequality-1*:
  assumes *I*: *finite I*
  assumes *ind*: *indep-vars* (λ *-. borel*) *X I*
  assumes *intsq*: $\bigwedge i.\ i \in I \Longrightarrow$ *integrable M* ($\lambda x.$ (*X i x*)^*2*)
  assumes *bnd*: $\bigwedge i.\ i \in I \Longrightarrow$ *AE x in M. X i x* $\leq$ *1*
  assumes *t*: $t \geq 0$
  defines $V \equiv$ ($\sum i \in I$. *expectation*($\lambda x.$ *X i x*^*2*))
  shows *prob* {$x \in$ *space M*. ($\sum i \in I$. *X i x* $-$ *expectation* (*X i*)) $\geq$
t} $\leq$
    exp ($-V *$ *bennett-h* (*t* / *V*))
⟨*proof*⟩

lemma *real-AE-le-sum*:
  assumes $\bigwedge i.\ i \in I \Longrightarrow$ *AE x in M. f i x* $\leq$ (*g i x::real*)
  shows *AE x in M.* ($\sum i{\in}I$. *f i x*) $\leq$ ($\sum i{\in}I$. *g i x*)
⟨*proof*⟩

lemma *real-AE-eq-sum*:
  assumes $\bigwedge i.\ i \in I \Longrightarrow$ *AE x in M. f i x* $=$ (*g i x::real*)
  shows *AE x in M.* ($\sum i{\in}I$. *f i x*) $=$ ($\sum i{\in}I$. *g i x*)
⟨*proof*⟩

theorem *bennett-inequality*:
  assumes *I*: *finite I*
  assumes *ind*: *indep-vars* (λ *-. borel*) *X I*
  assumes *intsq*: $\bigwedge i.\ i \in I \Longrightarrow$ *integrable M* ($\lambda x.$ (*X i x*)^*2*)
  assumes *bnd*: $\bigwedge i.\ i \in I \Longrightarrow$ *AE x in M. X i x* $\leq B$
  assumes *t*: $t \geq 0$
  assumes *B*: $B > 0$
  defines $V \equiv$ ($\sum i \in I$. *expectation* ($\lambda x.$ *X i x*^*2*))
  shows *prob* {$x \in$ *space M*. ($\sum i \in I$. *X i x* $-$ *expectation* (*X i*)) $\geq$
t} $\leq$
    exp ($-$ *V* / *B*^*2* $*$ *bennett-h* (*t* $*$ *B* / *V*))
⟨*proof*⟩

lemma *bennett-h-bernstein-bound*:
  assumes $x \geq 0$
  shows *bennett-h x* $\geq$ *x*^*2* / (*2* $*$ (*1* + *x* / *3*))
⟨*proof*⟩

lemma *sum-sq-exp-eq-zero-imp-zero*:
  assumes *finite I i* $\in$ *I*
  assumes *intsq*: *integrable M* ($\lambda x.$ (*X i x*)^*2*)
  assumes ($\sum i \in I$. *expectation* ($\lambda x.$ *X i x*^*2*)) $=$ *0*

shows *AE x in M. X i x* = (*0*::*real*)
⟨*proof*⟩

corollary *bernstein-inequality*:
assumes *I*: *finite I*
assumes *ind*: *indep-vars* (λ -. *borel*) *X I*
assumes *intsq*: $\bigwedge i.\ i \in I \Longrightarrow$ *integrable M* (λ*x*. (*X i x*)^*2*)
assumes *bnd*: $\bigwedge i.\ i \in I \Longrightarrow$ *AE x in M. X i x* ≤ *B*
assumes *t*: *t* ≥ *0*
assumes *B*: *B* > *0*
defines *V* ≡ (∑ *i* ∈ *I*. *expectation* (λ*x*. *X i x*^*2*))
shows *prob* {*x* ∈ *space M*. (∑ *i* ∈ *I*. *X i x* − *expectation* (*X i*)) ≥ *t*} ≤
exp (− (*t*^*2* / (*2* ∗ (*V* + *t* ∗ *B* / *3*))))
⟨*proof*⟩

end

end

3 Bienaymé's identity

Bienaymé's identity [5, §17] can be used to deduce the variance of a sum of random variables, if their co-variance is known. A common use-case of the identity is the computation of the variance of the mean of pair-wise independent variables.

theory *Bienaymes-Identity*
imports *Concentration-Inequalities-Preliminary*
begin

context *prob-space*
begin

lemma *variance-divide*:
fixes *f* :: ′*a* ⇒ *real*
assumes *integrable M f*
shows *variance* (λω. *f* ω / *r*) = *variance f* / *r*^*2*
⟨*proof*⟩

definition *covariance* **where**
covariance f g = *expectation* (λω. (*f* ω − *expectation f*) ∗ (*g* ω − *expectation g*))

lemma *covariance-eq*:
fixes *f* :: ′*a* ⇒ *real*
assumes *f* ∈ *borel-measurable M g* ∈ *borel-measurable M*
assumes *integrable M* (λω. *f* ω^*2*) *integrable M* (λω. *g* ω^*2*)
shows *covariance f g* = *expectation* (λω. *f* ω ∗ *g* ω) − *expectation f*

$*$ *expectation g*
⟨*proof*⟩

lemma *covar-integrable*:
  fixes $f\ g$ :: $'a \Rightarrow real$
  assumes $f \in$ *borel-measurable* M $g \in$ *borel-measurable* M
  assumes *integrable* M $(\lambda\omega.\ f\ \omega\hat{}2)$ *integrable* M $(\lambda\omega.\ g\ \omega\hat{}2)$
  shows *integrable* M $(\lambda\omega.\ (f\ \omega - \mathit{expectation}\ f) * (g\ \omega - \mathit{expectation}\ g))$
⟨*proof*⟩

lemma *sum-square-int*:
  fixes f :: $'b \Rightarrow 'a \Rightarrow real$
  assumes *finite* I
  assumes $\bigwedge i.\ i \in I \Longrightarrow f\ i \in$ *borel-measurable* M
  assumes $\bigwedge i.\ i \in I \Longrightarrow$ *integrable* M $(\lambda\omega.\ f\ i\ \omega\hat{}2)$
  shows *integrable* M $(\lambda\omega.\ (\sum i \in I.\ f\ i\ \omega)^2)$
⟨*proof*⟩

theorem *bienaymes-identity*:
  fixes f :: $'b \Rightarrow 'a \Rightarrow real$
  assumes *finite* I
  assumes $\bigwedge i.\ i \in I \Longrightarrow f\ i \in$ *borel-measurable* M
  assumes $\bigwedge i.\ i \in I \Longrightarrow$ *integrable* M $(\lambda\omega.\ f\ i\ \omega\hat{}2)$
  shows
    variance $(\lambda\omega.\ (\sum i \in I.\ f\ i\ \omega)) = (\sum i \in I.\ (\sum j \in I.\ \mathit{covariance}\ (f\ i)\ (f\ j)))$
⟨*proof*⟩

lemma *covar-self-eq*:
  fixes f :: $'a \Rightarrow real$
  shows *covariance* $f\ f$ = *variance* f
  ⟨*proof*⟩

lemma *covar-indep-eq-zero*:
  fixes $f\ g$ :: $'a \Rightarrow real$
  assumes *integrable* $M\ f$
  assumes *integrable* $M\ g$
  assumes *indep-var borel f borel g*
  shows *covariance* $f\ g = 0$
⟨*proof*⟩

lemma *bienaymes-identity-2*:
  fixes f :: $'b \Rightarrow 'a \Rightarrow real$
  assumes *finite* I
  assumes $\bigwedge i.\ i \in I \Longrightarrow f\ i \in$ *borel-measurable* M
  assumes $\bigwedge i.\ i \in I \Longrightarrow$ *integrable* M $(\lambda\omega.\ f\ i\ \omega\hat{}2)$
  shows *variance* $(\lambda\omega.\ (\sum i \in I.\ f\ i\ \omega)) =$
    $(\sum i \in I.\ \mathit{variance}\ (f\ i)) + (\sum i \in I.\ \sum j \in I - \{i\}.\ \mathit{covariance}$

$(f\ i)\ (f\ j))$
⟨*proof*⟩

theorem *bienaymes-identity-pairwise-indep*:
fixes $f :: 'b \Rightarrow 'a \Rightarrow real$
assumes *finite I*
assumes $\bigwedge i.\ i \in I \Longrightarrow f\ i \in borel\text{-}measurable\ M$
assumes $\bigwedge i.\ i \in I \Longrightarrow integrable\ M\ (\lambda\omega.\ f\ i\ \omega\hat{\ }2)$
assumes $\bigwedge i\ j.\ i \in I \Longrightarrow j \in I \Longrightarrow i \neq j \Longrightarrow indep\text{-}var\ borel\ (f\ i)$
$borel\ (f\ j)$
shows $variance\ (\lambda\omega.\ (\sum i \in I.\ f\ i\ \omega)) = (\sum i \in I.\ variance\ (f\ i))$
⟨*proof*⟩

lemma *indep-var-from-indep-vars*:
assumes $i \neq j$
assumes $indep\text{-}vars\ (\lambda\text{-}.\ M')\ f\ \{i, j\}$
shows $indep\text{-}var\ M'\ (f\ i)\ M'\ (f\ j)$
⟨*proof*⟩

lemma *bienaymes-identity-pairwise-indep-2*:
fixes $f :: 'b \Rightarrow 'a \Rightarrow real$
assumes *finite I*
assumes $\bigwedge i.\ i \in I \Longrightarrow f\ i \in borel\text{-}measurable\ M$
assumes $\bigwedge i.\ i \in I \Longrightarrow integrable\ M\ (\lambda\omega.\ f\ i\ \omega\hat{\ }2)$
assumes $\bigwedge J.\ J \subseteq I \Longrightarrow card\ J = 2 \Longrightarrow indep\text{-}vars\ (\lambda\ \text{-}.\ borel)\ f\ J$
shows $variance\ (\lambda\omega.\ (\sum i \in I.\ f\ i\ \omega)) = (\sum i \in I.\ variance\ (f\ i))$
⟨*proof*⟩

lemma *bienaymes-identity-full-indep*:
fixes $f :: 'b \Rightarrow 'a \Rightarrow real$
assumes *finite I*
assumes $\bigwedge i.\ i \in I \Longrightarrow f\ i \in borel\text{-}measurable\ M$
assumes $\bigwedge i.\ i \in I \Longrightarrow integrable\ M\ (\lambda\omega.\ f\ i\ \omega\hat{\ }2)$
assumes $indep\text{-}vars\ (\lambda\ \text{-}.\ borel)\ f\ I$
shows $variance\ (\lambda\omega.\ (\sum i \in I.\ f\ i\ \omega)) = (\sum i \in I.\ variance\ (f\ i))$
⟨*proof*⟩

end

end

4 Cantelli's Inequality

Cantelli's inequality [3] is an improvement of Chebyshev's inequality for one-sided tail bounds.

theory *Cantelli-Inequality*
imports *HOL−Probability.Probability*
begin

context *prob-space*
begin

lemma *cantelli-arith*:
  assumes $a > (0::real)$
  shows $(V + (V / a)\hat{}2) / (a + (V / a))\hat{}2 = V / (a \hat{}2 + V)$ (**is** $?L = ?R$)
$\langle proof \rangle$

theorem *cantelli-inequality*:
  assumes [*measurable*]: *random-variable borel Z*
  assumes *intZsq*: *integrable M* $(\lambda z.\ Z\ z\hat{}2)$
  assumes *a*: $a > 0$
  shows *prob* $\{z \in space\ M.\ Z\ z - expectation\ Z \geq a\} \leq$
    variance Z $/ (a\hat{}2 +$ *variance Z*$)$
$\langle proof \rangle$

corollary *cantelli-inequality-neg*:
  assumes [*measurable*]: *random-variable borel Z*
  assumes *intZsq*: *integrable M* $(\lambda z.\ Z\ z\hat{}2)$
  assumes *a*: $a > 0$
  shows *prob* $\{z \in space\ M.\ Z\ z - expectation\ Z \leq -a\} \leq$
    variance Z $/ (a\hat{}2 +$ *variance Z*$)$
$\langle proof \rangle$

end

end

5 Efron-Stein Inequality

In this section we verify the Efron-Stein inequality. The verified theorem is stated as Efron-Stein inequality for non-symmetric functions by Steele [8]. However most textbook refer to this version as "the Efron-Stein inequality". The original result that was shown by Efron and Stein is a tail bound for the variance of a symmetric functions of i.i.d. random variables [4].

theory *Efron-Stein-Inequality*
  imports *Concentration-Inequalities-Preliminary*
begin

theorem *efron-stein-inequality-distr*:
  fixes f :: - $\Rightarrow$ *real*
  assumes *finite I*
  assumes $\bigwedge i.\ i \in I \Longrightarrow$ *prob-space* $(M\ i)$

assumes *integrable* $(PiM\ I\ M)$ $(\lambda x.\ f\ x\hat{}2)$ **and** *f-meas*: $f \in$ *borel-measurable* $(PiM\ I\ M)$
shows *prob-space.variance* $(PiM\ I\ M)\ f \leq$
$(\sum i{\in}I.\ (\int x.\ (f\ (\lambda j.\ x\ (j,False)) - f\ (\lambda j.\ x\ (j,\ j{=}i)))\hat{}2\ \partial PiM\ (I{\times}UNIV)\ (M \circ fst)))\ /\ 2$
(**is** $?L \leq ?R$)
⟨*proof*⟩

theorem (**in** *prob-space*) *efron-stein-inequality-classic*:
fixes f :: - $\Rightarrow$ *real*
assumes *finite I*
assumes *indep-vars* $(M' \circ fst)\ X\ (I \times (UNIV :: bool\ set))$
assumes $f \in$ *borel-measurable* $(PiM\ I\ M')$
assumes *integrable* M $(\lambda\omega.\ f\ (\lambda i{\in}I.\ X\ (i,False)\ \omega)\hat{}2)$
assumes $\bigwedge i.\ i \in I \Longrightarrow distr\ M\ (M'\ i)\ (X\ (i,\ True)) = distr\ M\ (M'\ i)\ (X\ (i,\ False))$
shows *variance* $(\lambda\omega.\ f\ (\lambda i{\in}I.\ X\ (i,False)\ \omega)) \leq$
$(\sum j \in I.\ expectation\ (\lambda\omega.\ (f\ (\lambda i{\in}I.\ X\ (i,False)\ \omega) - f\ (\lambda i{\in}I.\ X\ (i,\ i{=}j)\ \omega))\hat{}2))/2$
(**is** $?L \leq ?R$)
⟨*proof*⟩

end

6 McDiarmid's inequality

In this section we verify McDiarmid's inequality [6, Lemma 1.2]. In the source and also further sources sometimes refer to the result as the "independent bounded differences" inequality.

theory *McDiarmid-Inequality*
imports *Concentration-Inequalities-Preliminary*
begin

lemma *Collect-restr-cong*:
assumes $A = B$
assumes $\bigwedge x.\ x \in A \Longrightarrow P\ x = Q\ x$
shows $\{x \in A.\ P\ x\} = \{x \in B.\ Q\ x\}$
⟨*proof*⟩

lemma *ineq-chain*:
fixes h :: *nat* $\Rightarrow$ *real*
assumes $\bigwedge i.\ i < n \Longrightarrow h\ (i{+}1) \leq h\ i$
shows $h\ n \leq h\ 0$
⟨*proof*⟩

lemma *restrict-subset-eq*:
assumes $A \subseteq B$
assumes *restrict f B = restrict g B*

shows *restrict f A = restrict g A*
⟨*proof*⟩

Bochner Integral version of Hoeffding's Lemma using *interval-bounded-random-variable.Hoeffdin*

lemma (**in** *prob-space*) *Hoeffdings-lemma-bochner*:
assumes $l > 0$ **and** *E0*: *expectation f* = *0*
assumes *random-variable borel f*
assumes *AE x in M. f x* ∈ {*a..b::real*}
shows *expectation* $(\lambda x.\ exp\ (l * f\ x)) \leq exp\ (l^2 * (b - a)^2\ /\ 8)$ (**is** *?L* ≤ *?R*)
⟨*proof*⟩

lemma (**in** *prob-space*) *Hoeffdings-lemma-bochner-2*:
assumes $l > 0$ **and** *E0*: *expectation f* = *0*
assumes *random-variable borel f*
assumes $\bigwedge x\ y.\ \{x,y\} \subseteq space\ M \Longrightarrow |f\ x - f\ y| \leq (c::real)$
shows *expectation* $(\lambda x.\ exp\ (l * f\ x)) \leq exp\ (l\hat{}2 * c\hat{}2\ /\ 8)$ (**is** *?L* ≤ *?R*)
⟨*proof*⟩

lemma (**in** *prob-space*) *Hoeffdings-lemma-bochner-3*:
assumes *expectation f* = *0*
assumes *random-variable borel f*
assumes $\bigwedge x\ y.\ \{x,y\} \subseteq space\ M \Longrightarrow |f\ x - f\ y| \leq (c::real)$
shows *expectation* $(\lambda x.\ exp\ (l * f\ x)) \leq exp\ (l\hat{}2 * c\hat{}2\ /\ 8)$ (**is** *?L* ≤ *?R*)
⟨*proof*⟩

Version of *product-sigma-finite.product-integral-singleton* without the condition that $M\ i$ has to be sigma finite for all i:

lemma *product-integral-singleton*:
fixes *f* :: - ⇒ -::{*banach*, *second-countable-topology*}
assumes *sigma-finite-measure* (*M i*)
assumes *f* ∈ *borel-measurable* (*M i*)
shows $(\int x.\ f\ (x\ i)\ \partial(PiM\ \{i\}\ M)) = (\int x.\ f\ x\ \partial(M\ i))$ (**is** *?L* = *?R*)
⟨*proof*⟩

Version of *product-sigma-finite.product-integral-fold* without the condition that $M\ i$ has to be sigma finite for all i:

lemma *product-integral-fold*:
fixes *f* :: - ⇒ -::{*banach*, *second-countable-topology*}
assumes $\bigwedge i.\ i \in I \cup J \Longrightarrow$ *sigma-finite-measure* (*M i*)
assumes $I \cap J = \{\}$
assumes *finite I*
assumes *finite J*
assumes *integrable* (*PiM* (*I* ∪ *J*) *M*) *f*
shows $(\int x.\ f\ x\ \partial PiM\ (I \cup J)\ M) = (\int x.\ (\int y.\ f\ (merge\ I\ J(x,y))\ \partial PiM\ J\ M)\ \partial PiM\ I\ M)$ (**is** *?L* = *?R*)

and *integrable* $(PiM\ I\ M)$ $(\lambda x.\ (\int y.\ f\ (merge\ I\ J(x,y))\ \partial PiM\ J\ M))$ (**is** *?I*)
and *AE x in PiM I M. integrable* $(PiM\ J\ M)$ $(\lambda y.\ f\ (merge\ I\ J(x,y)))$ (**is** *?T*)
⟨*proof*⟩

lemma *product-integral-insert*:
fixes f :: - $\Rightarrow$ -::{*banach*, *second-countable-topology*}
assumes $\bigwedge k.\ k \in \{i\} \cup J \Longrightarrow$ *sigma-finite-measure* $(M\ k)$
assumes $i \notin J$
assumes *finite J*
assumes *integrable* $(PiM\ (insert\ i\ J)\ M)\ f$
shows $(\int x.\ f\ x\ \partial PiM\ (insert\ i\ J)\ M) = (\int x.\ (\int y.\ f\ (y(i := x))\ \partial PiM\ J\ M)\ \partial M\ i)$ (**is** *?L = ?R*)
⟨*proof*⟩

lemma *product-integral-insert-rev*:
fixes f :: - $\Rightarrow$ -::{*banach*, *second-countable-topology*}
assumes $\bigwedge k.\ k \in \{i\} \cup J \Longrightarrow$ *sigma-finite-measure* $(M\ k)$
assumes $i \notin J$
assumes *finite J*
assumes *integrable* $(PiM\ (insert\ i\ J)\ M)\ f$
shows $(\int x.\ f\ x\ \partial PiM\ (insert\ i\ J)\ M) = (\int y.\ (\int x.\ f\ (y(i := x))\ \partial M\ i)\ \partial PiM\ J\ M)$ (**is** *?L = ?R*)
⟨*proof*⟩

lemma *merge-empty*[*simp*]:
merge {} I (y,x) = *restrict* x I
merge I {} (y,x) = *restrict* y I
⟨*proof*⟩

lemma *merge-cong*:
assumes *restrict x1 I = restrict x2 I*
assumes *restrict y1 J = restrict y2 J*
shows *merge I J (x1,y1) = merge I J (x2,y2)*
⟨*proof*⟩

lemma *restrict-merge*:
restrict (*merge I J x*) K = *merge* $(I \cap K)$ $(J \cap K)$ x
⟨*proof*⟩

lemma *map-prod-measurable*:
assumes $f \in M \rightarrow_M M'$
assumes $g \in N \rightarrow_M N'$
shows *map-prod f g* $\in M \bigotimes_M N \rightarrow_M M' \bigotimes_M N'$
⟨*proof*⟩

lemma *mc-diarmid-inequality-aux*:
fixes f :: $(nat \Rightarrow 'a) \Rightarrow real$

fixes n :: *nat*
assumes $\bigwedge i.\ i < n \Longrightarrow$ *prob-space* $(M\ i)$
assumes $\bigwedge i\ x\ y.\ i{<}n \Longrightarrow \{x,y\} \subseteq$ *space* $(PiM\ \{..{<}n\}\ M) \Longrightarrow (\forall j{\in}\{..{<}n\}{-}\{i\}.\ x\ j{=}y\ j) \Longrightarrow |f\ x{-}f\ y| {\le} c\ i$
assumes *f-meas*: $f \in$ *borel-measurable* $(PiM\ \{..{<}n\}\ M)$ **and** ε*-gt-0*: $\varepsilon > 0$
shows $\mathcal{P}(\omega\ in\ PiM\ \{..{<}n\}\ M.\ f\ \omega - (\int \xi.\ f\ \xi\ \partial PiM\ \{..{<}n\}\ M) \ge \varepsilon) \le exp\ (-(2{*}\varepsilon\hat{\ }2)/(\sum i{<}n.\ (c\ i)\hat{\ }2))$
(**is** *?L* $\le$ *?R*)
⟨*proof*⟩

theorem *mc-diarmid-inequality-distr*:
fixes f :: $('i \Rightarrow 'a) \Rightarrow$ *real*
assumes *finite* I
assumes $\bigwedge i.\ i \in I \Longrightarrow$ *prob-space* $(M\ i)$
assumes $\bigwedge i\ x\ y.\ i \in I \Longrightarrow \{x,y\} \subseteq$ *space* $(PiM\ I\ M) \Longrightarrow (\forall j{\in}I{-}\{i\}.\ x\ j{=}y\ j) \Longrightarrow |f\ x{-}f\ y| {\le} c\ i$
assumes *f-meas*: $f \in$ *borel-measurable* $(PiM\ I\ M)$ **and** ε*-gt-0*: $\varepsilon > 0$
shows $\mathcal{P}(\omega\ in\ PiM\ I\ M.\ f\ \omega - (\int \xi.\ f\ \xi\ \partial PiM\ I\ M) \ge \varepsilon) \le exp\ (-(2{*}\varepsilon\hat{\ }2)/(\sum i{\in}I.\ (c\ i)\hat{\ }2))$
(**is** *?L* $\le$ *?R*)
⟨*proof*⟩

lemma (**in** *prob-space*) *mc-diarmid-inequality-classic*:
fixes f :: $('i \Rightarrow 'a) \Rightarrow$ *real*
assumes *finite* I
assumes *indep-vars* $N\ X\ I$
assumes $\bigwedge i\ x\ y.\ i \in I \Longrightarrow \{x,y\} \subseteq$ *space* $(PiM\ I\ N) \Longrightarrow (\forall j{\in}I{-}\{i\}.\ x\ j{=}y\ j) \Longrightarrow |f\ x{-}f\ y| {\le} c\ i$
assumes *f-meas*: $f \in$ *borel-measurable* $(PiM\ I\ N)$ **and** ε*-gt-0*: $\varepsilon > 0$
shows $\mathcal{P}(\omega\ in\ M.\ f\ (\lambda i{\in}I.\ X\ i\ \omega) - (\int \xi.\ f\ (\lambda i{\in}I.\ X\ i\ \xi)\ \partial M) \ge \varepsilon) \le exp\ (-(2{*}\varepsilon\hat{\ }2)/(\sum i{\in}I.\ (c\ i)\hat{\ }2))$
(**is** *?L* $\le$ *?R*)
⟨*proof*⟩

end

7 Paley-Zygmund Inequality

This section proves slight improvements of the Paley-Zygmund Inequality [7]. Unfortunately, the improvements are on Wikipedia with no citation.

theory *Paley-Zygmund-Inequality*
imports *Lp.Lp*
begin

context *prob-space*
begin

theorem *paley-zygmund-inequality-holder*:
  assumes *p*: $1 < (p{::}real)$
  assumes *rv*: *random-variable borel Z*
  assumes *intZp*: *integrable M* $(\lambda z.\ |Z\ z|\ powr\ p)$
  assumes *t*: $\vartheta \le 1$
  assumes *ZAEpos*: *AE z in M*. $Z\ z \ge 0$
  shows
    $(expectation\ (\lambda x.\ |Z\ x - \vartheta * expectation\ Z|\ powr\ p)\ powr\ (1\ /\ (p{-}1))) *$
    $prob\ \{z \in space\ M.\ Z\ z > \vartheta * expectation\ Z\}$
      $\ge ((1{-}\vartheta)\ powr\ (p\ /\ (p{-}1)) * expectation\ Z\ powr\ (p\ /\ (p{-}1)))$
⟨*proof*⟩

corollary *paley-zygmund-inequality*:
  assumes *rv*: *random-variable borel Z*
  assumes *intZsq*: *integrable M* $(\lambda z.\ (Z\ z)\hat{\ }2)$
  assumes *t*: $\vartheta \le 1$
  assumes *Zpos*: $\bigwedge z.\ z \in space\ M \Longrightarrow Z\ z \ge 0$
  shows
    $(variance\ Z + (1{-}\vartheta)\hat{\ }2 * (expectation\ Z)\hat{\ }2) *$
    $prob\ \{z \in space\ M.\ Z\ z > \vartheta * expectation\ Z\}$
      $\ge (1{-}\vartheta)\hat{\ }2 * (expectation\ Z)\hat{\ }2$
⟨*proof*⟩

end

end

References

[1] G. Bennett. Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association*, 57(297):33–45, 1962.

[2] S. Boucheron, G. Lugosi, and O. Bousquet. Concentration inequalities. In O. Bousquet, U. von Luxburg, and G. Rätsch, editors, *Advanced Lectures on Machine Learning, ML Summer Schools 2003, Canberra, Australia, February 2-14, 2003, Tübingen, Germany, August 4-16, 2003, Revised Lectures*, volume 3176 of *Lecture Notes in Computer Science*, pages 208–240. Springer, 2003.

[3] F. P. Cantelli. Sui confini della probabilita. In *Atti del Congresso Internazionale dei Matematici: Bologna del 3 al 10 de settembre di 1928*, pages 47–60, 1929.

[4] B. Efron and C. Stein. The Jackknife Estimate of Variance. *The Annals of Statistics*, 9(3):586 – 596, 1981.

[5] M. Loève. *Probability Theory I*, chapter Sums of Independent Random Variables, pages 235–279. Springer New York, New York, NY, 1977.

[6] C. McDiarmid. *Surveys in Combinatorics, 1989: Invited Papers at the Twelfth British Combinatorial Conference*, chapter On the method of bounded differences, pages 148 – 188. London Mathematical Society Lecture Note Series. Cambridge University Press, 1989.

[7] R. E. Paley and A. Zygmund. A note on analytic functions in the unit circle. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 28, pages 266–272. Cambridge University Press, 1932.

[8] J. M. Steele. An Efron-Stein Inequality for Nonsymmetric Statistics. *The Annals of Statistics*, 14(2):753 – 758, 1986.